COHEN

DECISION SCIENCE CONSORTIUM, INC.

AN EXPERT SYSTEM FRAMEWORK FOR ADAPTIVE EVIDENTIAL REASONING: APPLICATION TO IN-FLIGHT ROUTE RE-PLANNING

Marvin S. Cohen, Kathryn B. Laskey, James McIntyre,
Bryan Thompson

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, VA 22043

(703) 790-0510

Sponsored by:

Air Force Wright Aeronautical Laboratories
Aeronautical Systems Division (AFSC)
United States Air Force
Wright-Patterson AFB, Ohio 45433-6503

Contract No. F33615-85-C-1832
Project No. FY1175-85-01394

March 21, 1986

FINAL REPORT

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **21 MAR 1986**	2. REPORT TYPE	3. DATES COVERED **00-00-1986 to 00-00-1986**
4. TITLE AND SUBTITLE **An Expert System Framework for Adaptive Evidential Reasoning: Application to In-Flight Route Re-Planning**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Decision Sciencc Consortium, Inc,7700 Lcesburg Pike, Suitc 421,Falls Church,VA,22043**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**141**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

AN EXPERT SYSTEM FRAMEWORK FOR ADAPTIVE EVIDENTIAL REASONING: APPLICATION TO IN-FLIGHT ROUTE RE-PLANNING

Marvin S. Cohen, Kathryn B. Laskey, James McIntyre,
Bryan Thompson

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, VA 22043

(703) 790-0510

Sponsored by:

Air Force Wright Aeronautical Laboratories
Aeronautical Systems Division (AFSC)
United States Air Force
Wright-Patterson AFB, Ohio 45433-6503

Contract No. F33615-85-C-1832
Project No. FY1175-85-01394

March 21, 1986

APPENDIX B
DOD No. 85.1

U.S. DEPARTMENT OF DEFENSE

SMALL BUSINESS INNOVATION RESEARCH PROGRAM
PHASE 1 — FY 1985
PROJECT SUMMARY

Topic No. 20 Military Department/Agency USAF.

Name and Address of Proposer

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, VA 22043

Name and Title of Principal Investigator

Marvin S. Cohen, Vice President

Proposer's Title*

Technical Abstract* (Limit your abstract to 200 words with no classified or proprietary information/data.)

The successful introduction of AI technology into Air Force avionics has been hindered by the need for intelligent real-time reasoning with incomplete and often inconsistent data. The primary objective of the present research has been to explore the feasibility of new methods for inference in avionics expert systems, which address specific shortcomings in existing approaches and combine some of their distinct virtues. A variety of approaches have been critically reviewed: quantitative representations (Bayes, Shafer, and Zadeh), qualitative frameworks (e.g., Doyle, Toulmin, P. Cohen), and efforts to synthesize logic and probability (Nilsson, Sage). An innovative framework for expert system reasoning has been developed which combines quantitative manipulation of uncertainty (via Shaferian belief functions), a qualitative frame for representing an evidential argument, and a non-monotonic capability for revising probabilistic arguments when they lead to conflicting results. This framework, along with a personalized user interface, has been implemented in a small-scale demonstration system for in-flight responses to pop-up threats, the Adaptive Route Replanner (ARR). Results with ARR strongly confirm the feasibility of a system which reasons intelligently and flexibly in the face of uncertainty.

Anticipated Benefits/Potential Commercial Applications of the Research or Development

The effective dissmenation of expert system technology, both in Air Force avionics and in other fields, will depend on the flexible and intelligent manipulation of uncertainty. ARR provides an expert system building tool which fills this void; hence, it has the potential for significant impact on commercial, academic, and governmental markets for AI products, as well as in the specific area of Air Force route planning.

List a maximum of 8 Key Words that describe the Project.

Expert Systems, Uncertainty, Avionics, Human-Computer Interface, Command and Control, AI, Decision Aids, Adaptive Systems

*Nothing on this page is classified or proprietary information/data

ACKNOWLEDGMENTS

The work reported here was sponsored by the Air Force Wright Aeronautical Laboratories at Wright-Patterson Air Force Base under Contract No. F33615-85-C-1832, as Phase I of a Small Business Innovation Research contract. We are grateful to Gurdial S. Saini for his guidance throughout the project, and to Jerry Covert, Greg Gunsch, Joe Ghrayeb, Bill Mallet, and Keith Jackson for their helpful advice.

We also owe thanks to Nancy Poudrier and Gail Hurd for their exemplary efforts in the production of the report and preparation of materials for the demonstration system.

TABLE OF CONTENTS

1.0 INTRODUCTION

1.1 Scope and Structure of Report

The primary purpose of the present research has been to demonstrate the feasibility of designing intelligent systems with the capacity for adaptive, flexible reasoning in uncertain and changing environments. The overall goals of the project are (1) the development of innovative inference frameworks for reasoning in avionics environments characterized by high stakes, large volumes of complex and conflicting information, and the need for rapid response; and (2) to lay the groundwork for a more general understanding of the process of choosing and designing inference frameworks for avionics expert systems applications. The achievement of these objectives would have far-reaching implications for the successful application of artificial intelligence technology in both military and civilian contexts, allowing the development of systems which fully exploit, while significantly improving upon, human intelligent reasoning. The ultimate result should be improved performance, at comparatively little cost, of a wide range of combat systems.

This report details the contributions of the initial phase of this research. The remainder of this introductory section provides background on the problem and briefly summarizes specific objectives. Sections 2 through 5 report on the results. Section 2 is a critical review of alternative inference theories. The review highlights the shortcomings of current approaches in providing fully adequate representations of uncertainty, and fully adequate techniques for adaptively manipulating uncertain beliefs. Section 3 describes the main product of the effort, an innovative framework for expert system inference in uncertain domains. Section 4 describes the application of that inference framework to an Air Force combat environment in a small-scale prototype system for in-flight route replanning. Section 5 describes a complementary line of research on concepts for human-computer interaction, and discusses their application in the prototype system. Section 6 summarizes the work and briefly explores directions for the future.

1.2 Background on the Problem

In recent years techniques of artificial intelligence (AI) have been employed to replicate, or improve on, human reasoning in an increasing sphere of inference and decision-making tasks (Hayes-Roth et al., 1983; Buchanan and Duda, 1982). Expert systems have now been developed for medical diagnosis and treatment (e.g., Shortliffe, 1976), geological exploration (e.g., Duda et al., 1979), chemical analysis (Lindsay et al., 1980), military planning (Engleman et al., 1979), and other areas of specialized human skill.

Unfortunately, the introduction of AI technologies into real-time tactical environments has been relatively slow. Among the reasons for lack of progress are a set of technical obstacles arising largely from the complexity of the inference task in these problem domains: (1) Near-future avionics environments will be characterized by high stakes and increasing numbers of high performance threats both in the air and on the ground, heightening both the time pressure and the uncertainty under which systems must function. Avionics decision aids must support real-time decisions in rapidly changing environments, while utilizing data sets which are large, incomplete, unreliable, and often inconsistent. (2) In air-to-air and air-to-ground combat, conflicts between critical objectives occur frequently and must be resolved. E.g., the requirements to communicate with other units, to localize threats by means of active emissions, or to move into proximity to a potential target all may conflict with the goal of concealing one's presence and location. (3) New advances in avionics technology have often led to the introduction of "black boxes" which are poorly integrated with other hardware or software components, and whose displays and controls are incompatible. Future systems therefore will involve exchange of outputs among subsystems which utilize radically different methods of representation and inference, e.g., rule-based architectures for predicting threat capabilities versus mathematical or statistical techniques of signal analysis, and must utilize a common set of user-system dialogue procedures. (4) In these environments human abilities to deal with unanticipated events, or ill-defined concepts, may be crucial. To obtain true synergy between human and computer capabilities, the aid must permit a dynamic and flexible partitioning of reasoning task components between human and computer, be able to communicate both the degree of

confidence and the rationale behind its recommendations, and facilitate intelligent override at any point in the reasoning process.

In recent years, serious attention has turned in the AI/expert systems community to the problem of reasoning about uncertainty (e.g., Buchanan and Shortliffe, 1984); and tentative theoretical steps have been made toward flexible systems capable of adaptive learning (e.g., Michalski et al., 1983). Nevertheless, current expert systems technology has for the most part failed to capture the capability of many domain experts to respond adaptively and flexibly to conditions which violate the original assumptions, to create new methods of reasoning where required, and to develop new ways of organizing data and collecting information based on unanticipated events. An adaptive capability of this sort is required in order to build systems that address the challenges of the modern battlefield.

We believe that a significant opportunity for addressing these shortcomings exists in recent work on inexact reasoning in artificial intelligence and in statistics.

The design of methods for inexact reasoning has in the past several years moved from the background into the forefront of attention in expert system research, and in AI more generally. In addition to the *ad hoc* numeric methods developed in such early systems as MYCIN and PROSPECTOR, a variety of formally justified quantitative approaches are now being discussed and implemented. Among the most prominent are variants of Bayesian probability theory, belief functions (Shafer, 1976), and fuzzy set or possibility theory (Zadeh, 1965, 1972). Nonnumerical approaches to reasoning with incomplete information have also been developed, and are perhaps closer to the mainstream AI tradition of symbolic reasoning: e.g., non-monotonic logic (Doyle, 1979; *Artificial Intelligence* (special issue), 1980); and the theory of endorsements (Cohen, 1985). Although there have been a few attempts (e.g., Nilsson, 1984; Ginsberg, 1984; Cohen, 1985) to integrate the numeric and non-numeric traditions, for the most part they have remained separate.

Uncertainty calculi will eventually be judged by how successfully they contribute to a variety of expert system functions; for example: (1) deriving the uncertainty of a conclusion from uncertainty in data and rules across

potentially lengthy lines of reasoning; (2) combining different items of evidence or outputs of different analytical subsystems; (3) resolving conflicts between different lines of reasoning (e.g., by collecting more information or by revising assumptions); and (4) displaying conclusions, explanations, and measures of confidence to users in ways that are readily understood. In time-stressed environments additional functions may include: (5) efficient allocation of resources among different lines of reasoning or information collection options, (6) halting computations when results are "acceptable enough" in the light of prevailing time and resource constraints.

Unfortunately, no current systems effectively encompass these diverse capabilities. Moreover, there has as yet been little systematic investigation of the impact of alternative inference frameworks on expert system functions. Alternative frameworks differ in the concept of uncertainty they attempt to capture (e.g., chance, imprecision, or incompleteness of evidence) and the degree to which appropriate normative justifications have been achieved. They differ also in the demands they impose on experts for assessments, in the computational burden they impose on the system, and in the ease with which they represent distinctions and yield conclusions which are natural to a particular expert, user, or problem domain. Design choices, in short, must be multidimensional. But it is by no means clear how tradeoffs among these competing considerations should be resolved.

Perhaps more importantly, current expert systems have typically incorporated rather primitive knowledge representation schemes (e.g., a homogeneous collection of rules), and such systems have been unable to duplicate the adaptive, iterative model revision process practiced by human experts.

The present work represents an initial effort to address these technical challenges in the context of avionics expert systems. The ultimate objectives are to develop improved inference architectures for pilot decision aid applications and to develop a more general understanding of the process of choosing and designing inference frameworks for avionics expert system applications. Existing approaches to reasoning about uncertainty have been critically analyzed and an innovative inference framework has been developed which incorporates elements of a variety of existing approaches and which provides a unique capability for adaptive self-improving inference. The feasibility of

this concept has been demonstrated in a specific avionics application area: i.e., inflight route re-planning in the face of strategic pop-up threats. A prototype aid has been developed for demonstration purposes which illustrates both the inference framework and the user/computer interface, and which will serve as a foundation for continued research and development.

1.3 Summary of Specific Objectives of the Research

In sum, the specific objectives of the present research (described in detail in the following sections) were:

- to perform a critical review and evaluation of alternative inference frameworks (including Bayesian, Shaferian, fuzzy, non-monotonic), identifying strengths and weaknesses for use in expert systems designed for real-time tactical environments (Section 2);
- to develop improved inference frameworks for real-time tactical expert systems (Section 3);
- to develop concepts for human-computer interaction (Section 5);
- to implement the developments of Phase I in a small-scale prototype system in a selected avionics context (Sections 4 and 5).

The overriding aim of the present research effort was to demonstrate the feasibility of our inference framework. From the results reported below it is clear that this objective has been met. Section 6 discusses directions for future research, both theoretical and applied.

2.0 REVIEW AND CRITIQUE OF ALTERNATE INFERENCE FRAMEWORKS

2.1 Overview.

For purposes of this review, we divide inference theories into three general categories. The first category is that of quantitative theories for representing and manipulating uncertainty. Of these, Bayesian probability theory (e.g., Savage, 1954) has the longest and most distinguished history. Recently, a great deal of attention has been devoted to two newer numerical theories: Shafer's (1976) theory of belief functions and Zadeh's (1965) theory of fuzzy logic.

The second category of inference theories is a set of qualitative, non-numeric inference frameworks. Our discussion begins with a brief mention of classical logic. Based on classical logic is the theory of *non-monotonic logic* (Doyle 1979), which is an outgrowth of theorem-proving systems in artificial intelligence. Non-monotonic logic allows for provisional acceptance of uncertain premises, which may later be retracted when they lead to contradictory conclusions. Toulmin (1958) introduces a new theory of logic based on an analogy with jurisprudence rather than the abstract mathematics of classical logic. Paul Cohen's (1985) theory of endorsements is another outgrowth of the artificial intelligence tradition. Cohen's system represents uncertainty about a rule or conclusion by qualitative *endorsements*, which are propagated through inferences to conclusions.

The third category of inference mechanisms consists of systems attempting to synthesize logic and probability in some way. Two approaches are discussed: Lagomasino and Sage (1985) ostensibly base their theory on Toulmin's theory of logic, while Nilsson (1984) uses classical logic.

2.2 Numerical inference theories.

2.2.1 Bayesian probabilistic inference.

Using probability theory for inexact reasoning. Probability theory has become central to modern scientific culture. As such, it is the obvious calculus to consider for handling inexactness in expert systems. Its supporters in this role date back to the early work on probabilistic information processing (see Edwards, 1966) and earlier;

more recent contributors have been de Dombal (1973), in the field of medical decision making, and Schum (1980) in the intelligence field.

The application of probabilistic reasoning to rule-based expert systems is complex, but it can be illustrated with a simple example. Part of an expert system for avionics applications could be a threat classification system. A rule in such a system might be:

```
IF (OBSERVED SIGNAL HAS FEATURE X)
     THEN (THREAT IS SA-4) LR = 2.3)
```

where LR quantifies the impact of the evidence (the signal feature) on the hypothesis (that the threat is an SA-4). LR is a likelihood ratio, i.e., the probability of a signal with feature of type X given that the threat is an SA-4 divided by the probability of that signal feature given that it is not an SA-4. Satisfaction of the antecedent of this rule would lead to a process of Bayesian updating, in which the impact of the new evidence is combined with the prior odds of the hypothesis being true. Suppose H is the hypothesis that the object is an SA-4. Then Bayes' Theorem gives, in odds-likelihood form,

$$\frac{Pr[H|D]}{Pr[\overline{H}|D]} = \frac{Pr[D|H]}{Pr[D|\overline{H}]} \cdot \frac{Pr[H]}{Pr[\overline{H}]}$$

where D is the data that the signal has feature X, and H is the hypothesis that some other classification of the threat is appropriate. To carry out a simple analysis of this kind, three assessments are required, namely $Pr[D|H]$, $Pr[D|\overline{H}]$ and $Pr[H]$, i.e., the likelihoods and the prior probability.

Work on Bayesian approaches to inference has advanced from a simple one-step application of Bayes' rule to the elaboration in recent research of rather complex structures capable of capturing a wide diversity of human inference tasks and prescriptive intuitions (e.g., Schum, 1979, 1981). Bayesian techniques, for example, are able to accommodate a number of different ways that items of evidence can be related to one another with respect to a hypothesis (Schum and Martin, 1980): e.g., they may be contradictory (reporting and denying the same event), corroboratively redundant (reporting the same event), cumulatively redundant (reporting different events which

reduce one another's evidential impact), or non-redundant (reporting different events which enhance or do not change one another's evidential impact). In other, more complex cases of interdependence, Bayesian techniques capture the evidential impact of biases in an information source or non-independence of information source sensitivity with respect to what is being observed.

The following discussion highlights the major strengths and weaknesses of the Bayesian approach to uncertain reasoning in expert systems.

Feasibility: Quantity of inputs. When one attempts to use Bayesian probability theory on real inference problems, one quickly becomes aware of the complexity of the task. This complexity led Shortliffe (apparently) to construct his calculus of certainty factors as an alternative (see Shortliffe, 1976, Section 3.2). Schum (1980, p. 207) ends his advocacy of the Bayesian approach with a negative note: "...now we have other problems. I believe nobody realized how many ingredients there would be and how complex the judgments about these ingredients would be even in apparently simple cases." In all but the most trivial cases, a proper Bayesian analysis requires a great many conditional probabilities to be assessed. Schum presents the analysis of a fairly simple legal trial involving 7 pieces of evidence (Salmon's pills) and shows that at least 27 probability judgments are needed, even if all reasonable independence conditions hold. As well as requiring a very large number of probability assessments, the relations between them are difficult to organize, and the coherence of the total set of assessments is often difficult to determine.

Two important lines of defense for Bayesians are (a) that simplifying assumptions can always be made, e.g., equal prior probabilities, conditional independence of events; and (b) that variables which one does not care to deal with may be "integrated out," i.e., the resulting probabilities are regarded as marginal ("averages") with respect to possible values of the ignored variables. Thus, a Bayesian model may be created which is as simple as one likes.

Unfortunately, however, the situation is not quite as clear cut as this. "Simplifying assumptions" must in some sense be *judgments* (e.g., that priors are roughly equal, that events are conditionally independent). Otherwise, one sacrifices the validity of the Bayesian approach. As one Bayesian (Lindley,

1984) has put it, the Bayesian argument shows you the things you have to think about; so, think about them. From the Bayesian point of view, an argument which omits these factors is simply spurious. In the case of "integrating out" certain variables, no formal problem presents itself, since from a theoretical point of view the results with and without such variables should be the same. In actual fact, however, the difference in plausibility of the overall analysis can be very great (as we shall note below). Thus, although the required number of assessments may in fact be reduced by either of these means, the difficulty of the judgments required to do so may be considerable. Schum speaks of them as "exquisitely subtle".

A quite different approach, which we shall explore in greater detail below, is to regard simplifying strategies as assumptions whose validity is tested *implicitly* through their *use* in reasoning. If the outcome of using such assumptions is plausible, the burden of explicitly judging their validity is avoided.

A related tactic is to accept the Bayesian framework as, in principle, the correct way to handle uncertainty, and divert our research interests to approximations that are as close as possible to the Bayesian norm. Indeed, Shortliffe (1976, p. 164) originally saw certainty factors as a device in this direction. Shortliffe, however, did not explicitly derive his theory as a special case of the more general Bayesian model. Adams (1976) showed that assumptions necessary to derive Shortliffe's postulates in some cases do not exist, and in other cases are far more restrictive and implausible than the usual assumptions of equal priors and conditional independence. We shall return to this topic in the discussion of Shafer's theory (Section 2.2.2).

Computational tractability. There is no known, computationally tractable method for propagating uncertainties consistently through an arbitrary Bayesian network. Restrictions of some sort on the kind of model that is utilized are necessary. The only question (as in the previous discussion of inputs) is whether the restrictions will be plausible (i.e., define a meaningful, useful special case of Bayesian modeling) or *ad hoc*. PROSPECTOR adopted the latter approach. More recently, Pearl (1982) and Kim (1983) have explored the former. They show that independence assumptions make sense, and probabilities can be propagated by simple local computations, if the inferen-

tial network has (a) a causal interpretation, and (b) the form of a Chow tree (i.e., it lacks undirected cycles). Unfortunately, not all real problems will fit this special structure.

If validity is not to be sacrificed, computational tractability for a Bayesian system can be purchased only in special cases; and even then, only at the cost of complex and subtle judgments regarding interdependence among items of knowledge and the overall structure of the inferential argument. As we shall see, the situation is quite similar for Shaferian belief functions. For this reason, Shafer (1984a) has recently argued, the introduction of probability into expert systems appears to be inconsistent with the modularity of knowledge representations that up to now has been the most salient characteristic of such systems.

In Section 3 we shall return to some of these questions. We will propose that a careful use of qualitative reasoning, superimposed upon a probabilistic system, may reduce the requirement for experts (or users) to address issues of interdependence and model structure explicitly, and make such assessments easier when they are required, without undo compromise of validity. In Section 4, we describe a small-scale pilot implementation of an avionics expert system based on this approach.

Validity: Axiomatic derivation. Bayesian probability theory has a preeminent, though perhaps not conclusive, claim to validity among current proposals for the handling of uncertainty. De Finetti (1937/1964) showed that unless your beliefs conform to the rules of probability, a clever opponent could make you the victim of a "Dutch book," i.e., a set of gambles you would accept, but in which you lose regardless of the outcome of an uncertain state of affairs. More recently, Lindley (1982) has given a new derivation. Suppose that people are going to measure the uncertainty of events by some method, and we wish to know how good they are at doing so. If we devise a scoring system of *any* sort--as along as (a) the score is a joint function of the uncertainty measure and the event's truth or falsity, and (b) scores are additive across different events--then no matter what events actually occur, the best achievable score will always go to a form of Bayesian probability. Lindley concludes that "only probability is a sensible description of uncertainty."

A common objection to this sort of demonstration is that we are *not* in fact always (or usually) faced with a malicious adversary or, indeed, with a scoring system. But the point is not that we are, or should somehow presume that we are, always subjected to such peculiar circumstances. Even if we *never* encounter these conditions, other things being equal, a system which has the property of working well in them is more desirable (in *all* circumstances) than one which does not. It is plausible than an adequate system of uncertainty would guard against a Dutch book. It is plausible that such a system would score high *if* we ever chose to score it.

The more fundamental objection, in our view, is that while probability theory has been shown uniquely to possess a desirable property, but has not been shown to be *uniquely justified*. Other systems of uncertainty may have desirable properties that probability theory lacks. (In particular, alternative theories might deal more adequately with different kinds of uncertainty, such as incompleteness of evidence or imprecision. In this regard, note that De Finetti's and Lindley's arguments do not apply to systems which provide more than a single measure of uncertainty for each event, such as the upper and lower measures in Shafer's theory, or fuzzy probabilities in Zadeh's.)

Nonetheless, it seems incontrovertible to us that the existence of foundational arguments such as those described is a strong plus for Bayesian theory.

Plausibility of instances. The thrust of Bayesian analysis is to improve, rather than to replicate ordinary thinking. Bayesians argue that if one's ordinary intuitions are probabilistically incoherent, they ought to be changed. We might expect, nevertheless, that these revisions of belief would typically lead to judgments that are regarded as more plausible *after* reflection. In other words, the plausibility of the axioms should outweigh the initial plausibility of an incoherent set of judgments. In some cases, this seems true, e.g., most people who understand an explanation of the "gambler's fallacy" seem to accept that it is a fallacy; in other cases, perhaps, it is not true (e.g., Slovic and Tversky, 1974).

There is another issue here which is, we feel, more important. Even if revised (hence, coherent) beliefs are more plausible than unrevised, incoherent ones, all the credit cannot go to Bayesian theory. The reason is,

that the selection of a specific revision is not uniquely determined by the requirement of coherence. Consider, again, the example above of inferring the chance of H, i.e., that a particular threat is an SA-4, based on analysis of a signal A. Bayesian theory tells us only that our assessment of Pr[H] should be the same as Pr[H|A]Pr[A] + Pr[H|A]Pr[A], which is based on our assessments of Pr[H|A], Pr[A], and Pr[H|A]. The theory provides no guidance in the case where the two are not equal. Coherence by itself does not dictate that the result of an analysis is to be preferred to a direct judgment. We might choose to revise one or more of the assessments in the analysis, rather than to revise Pr[H].

This problem, which we may call the *incompleteness* of Bayesian theory, is exacerbated by the fact that in any problem there is more than one possible form of analysis. Many advocates and many critics of the Bayesian approach seem to imply that there is only one way a probabilistic analysis could be carried out and only one possible conclusion. To see that this is not the case, we return to the example of inferring H. Let B be intelligence information that the country in question had purchased in the last year an important component required for construction of an SA-4 installation. Instead of, or in addition to, conditioning our assessment on A, as above, we could condition on B, namely

$$\Pr[H] = \Pr[H|B]\Pr[B] + \Pr[H|\bar{B}]\Pr[\bar{B}].$$

Yet again, we could condition jointly on A and B:

$$\Pr[H] = \Pr[H|AB]\Pr[AB] + \Pr[H|A\bar{B}]\Pr[A\bar{B}] + \Pr[H|\bar{A}B]\Pr[\bar{A}B] + \Pr[H|\overline{AB}]\Pr[\overline{AB}].$$

Still more choices are open to us: for example, we could assess Pr[AB] directly, and/or further analyze it as Pr[A|B]Pr[B], and/or as Pr[B|A]Pr[A].

The Bayesian *theoretical* attitude is straightforward, namely that it does not matter which of these forms of analysis we perform or which answer we select, since coherent probability assessors should derive the same number whichever method they choose. Theory, however, is of <u>use</u> because we are not ordinarily coherent in our assessments. An analysis may well give us a different estimate of Pr[H] than if we directly judged it; otherwise, we wouldn't bother

with the analysis. Moreover, different analyses may well give us different answers; otherwise, we would have no cause for regarding some analyses as "better" than others.

An important assumption of Bayesian theory is that all analyses (by the same person) are based on the same evidence; they do not differ in the knowledge they draw upon. We would argue that this is, psychologically, not true. Different ways of formulating the same problem may well tap different internal stores of information. What is missing from the Bayesian framework is some notion of the quality of probability inputs, i.e., the amount of knowledge or completeness of evidence that underlies them. Several points can be made:

- Revision of probability judgments should be guided by a judgment of their quality, i.e., the amount of knowledge they represent.
- More than one analysis may be of value, if they bring different knowledge to bear on a problem (cf., Brown and Lindley, 1982, 1985).
- The application of Bayesian theory to a problem is not necessarily a linear process in which inputs are provided and conclusions computed. It is (or often should be) an iterative process, in which comparison of conclusions arrived at by different methods leads to revisions of inputs and assumptions, until overall consistency is achieved.

In ordinary statistical problem solving, perhaps, judgments of quality may safely remain implicit. But a major limitation in the automation of Bayesian theory within expert systems is the lack of an *explicit* measure of completeness of evidence, and a mechanism for its use in the revision of probability estimates.

This will be a major focus in our discussion of Shafer, below, and in the new developments to be described in Section 3.

Semantics: Behavioral specification. Bayesian theory provides a clear behavioral interpretation of probabilities in terms of preferences among bets. We can know what someone's probabilistic beliefs are by observing their actions under specified conditions. By contrast, a common complaint by Bayesians regarding other theories is the difficulty of knowing what the basic measures mean.

There are three different, but related, misunderstandings of this "operational definition." First, critics point out that betting may be an awkward and in some cases an impossible method for eliciting probabilities. It is often easier to ask for direct verbal judgments. There is a standard answer to this point by sophisticated Bayesians: Meaning need not be equated with evidence. Bayesians can use any method they like for estimating your probabilities, if there is a reasonable expectation that the result will match, or at least approximate, what they would have gotten had they used the betting paradigm.

This response hides a more subtle misunderstanding. It is still assumed that we can, at least in principle, always know what a person's probabilities are, simply by testing his preferences among bets. Since the operational definition specifies a situation where he must make a choice, it is implied that any person "has" probabilities waiting to be uncovered or "elicited". Is Bayesianism thus inevitable? This conception seems to be contradicted by the incoherence we typically find in people's unaided judgments, and which is amply documented in the experimental psychology literature (e.g., Kahneman, Slovic, and Tversky, 1982).

The sophisticated Bayesian was right, we suggest, in distinguishing meaning and evidence. But--sophisticated as he is--he has not absorbed the full implications of that distinction. Although he permits other kinds of evidence, he is still equating meaning with a particular observable operation. The problem, as pointed out by Quine (1953) and others in a more general critique of positivism, is that the selection of this rather than some other component of the theory as a "definition" is arbitrary. To return to our earlier example, suppose we equate Pr[H] for a person X with X's betting behavior in regard to H. Then we determine in the same way his value for Pr[H|A], Pr[H|A], and Pr[A]. Finally, we compute a new probability of H, Pr'[H], from the latter three values. Why shouldn't we define X's probability for H in terms of *this* operation, i.e., as Pr'[H]? One reply is that this operation requires a theoretical assumption viz., that X is coherent, to justify the computation of Pr'[H] from Pr[H|A], Pr[H|$\bar{A}$], and Pr[A]. But the earlier "operational definition" *could* be regarded as theoretical, too, since it is a theoretical hypothesis (i.e., that X acts so as to maximize subjectively expected utility) that enables us to derive X's probability for H from his

preferences among gambles involving H. Conversely, we *could* regard the definition in terms of Pr'[H] as purely "behavioral", by ignoring the theoretical hypotheses implicit in our calculations.

It is far more natural to regard *all* these potential "definitions" simply as theoretical predictions. How then, without definitions, do we assess the probabilities and utilities required to derive the predictions? The answer is that testing a theory is, inevitably, a bootstrapping operation, in which we use the theory, as if it were true, to estimate values for an interrelated set of parameters, then test for consistency of the results. If the results are consistent, the theory is confirmed; if not, it is disconfirmed. (For a general discussion see Glymore, 1980.) To the extent that people are probabilistically incoherent, therefore, probability theory is disconfirmed, and they cannot be regarded as "having" probabilities at all.

Have we overlooked the difference between descriptive and prescriptive theories? Perhaps "operational definitions" make sense for probabilities because they form part of a prescriptive theory. On the contrary, we suggest that there is a strong and important parallel between theory testing, as we just described it, and prescriptive analysis. Just as in descriptive science, we assume the prescriptive theory to be true, use it to perform a set of interrelated analyses, and then test them for consistency. However, if we find inconsistency among alternative prescriptive analyses, or between an analysis and direct judgment, we do not (necessarily) drop the prescriptive theory; we may choose to revise the values in one or more analyses so as to *make* them consistent. In so doing, we *construct* rather than discover or confirm a probability model for our beliefs.

What then is left of the Bayesian claim that operational definitions are required for clarity of concepts? The third and final misunderstanding we wish to address is the notion that because "operational definitions" are arbitrary, and do not guarantee the applicability or even the relevance of a prescriptive theory, that *behavioral specification* is of no use. In fact, it is quite critical: without it, there is no link, or else no clear link, between the prescriptive theory and action. With it, the prescriptive process described above, in which a coherent set of judgments is arrived at through successive iterations, also produces a clear set of implications for action.

In expert system applications, such implications are typically the reason for developing the system. Moreover, such specifications may play a clarifying role for the decision maker in the process of iteratively arriving at an appropriate set of judgments. The existence of such specifications must, therefore, be counted as a plus for the Bayesian theory.

Naturalness of inputs. Behavioral specification is not sufficient to guarantee the usefulness of an inference framework. A common objection to Bayesian theory urged by proponents of alternative views, is that the inputs it requires exceed, in various ways, the capabilities of the decision makers it is designed to aid. A distinction must be made between two types of claim against Bayesian theory: that it fails adequately to deal with *imprecision* and with *incompleteness of evidence*.

Bayesians assume that experts are capable of quantifying their uncertainties and values to an arbitrary degree of precision. But this is true of no other known process of measurement. Experts may simply not know, to the required exactitude, what their beliefs or preferences are.

Alternately, the evidence may be *incomplete* in that it does not justify the degree of confidence suggested by use of a single number to assess an uncertainty. Some assessments (e.g., the probability that the Soviets will invade Western Europe within the next year) are less well supported than others (e.g., the probability that a coin in my pocket will land heads if tossed). In the former cases, the available evidence may justify no more than a range of probabilities rather than a single number.

There is an important distinction between these two complaints: the first is consistent with the basic prescriptive adequacy of probability theory, but seeks to accommodate human shortcomings in the assessment task. In contrast, the second objection has a normative basis: probabilities themselves are inappropriate where evidence is incomplete. We shall explore these positions in more detail in our discussions of Zadeh and Shafer, respectively.

A related problem is that the Bayesian framework addresses *probabilistic* and not *causal* relationships. In many instances (particularly for applications for which rule-based expert systems are suited) people's reasoning processes

are naturally causally oriented (Abelson, 1985; Ross, 1977; Tversky and Kahneman, 1982). People may interpret probabilistic information causally, leading to commonly observed biases such as ignoring base rates or the conjunction fallacy. The persistence of such biases (Tversky and Kahneman, 1983) points to the difficulty of translating causal reasoning into probabilistic judgments.

Concepts of uncertainty. Bayesian theory is clearly designed to capture the concept of chance, or uncertainty about facts. We argued that an important gap in Bayesian theory is the lack of a measure of completeness or quality of evidence, i.e., the lack of a distinction between firm probabilities (.5 as the probability of heads on a coin toss) and those based on guesswork (.5 as the probability of a Soviet invasion). Intuitively, the weight of evidence supporting some probability judgments is stronger that that supporting others. We argued that this concept in fact plays an important role in ordinary applications of probability theory, by guiding the choice among potential revisions of belief in the light of an analysis or set of analyses. We hope to demonstrate below (Section 3) that an *explicit* measure of this sort is critical for the control of reasoning in an expert system that intelligently handles uncertainty about facts.

To what extent could Bayesian theory itself be extended to cover the concept of completeness of evidence? Lindley et al. (1979) have recently attempted to formalize the intuitive notion that we are firmer about some probability assessments than others. The tool they introduce is a second-order probability distribution over possible values of the true first-order probability. The spread of the second-order distribution is a measure of the firmness of the original probabilities. Lindley et al. have described procedures for statistically aggregating inconsistent probabilistic analyses by means of such second-order judgments.

These efforts have failed, in our opinion, for a variety of reasons. *Feasibility:* The quantity and difficulty of required inputs is increased, rather than decreased, to the degree that one's evidence is incomplete. Computational intractability will certainly be increased as well. *Validity:* Axiomatic justifications and behavioral specifications which apply to first-order probabilities become much less convincing at higher levels, where, for

example, gambles or scores which depend on one's own "true" probabilities, rather than actual events, lack plausibility. Face validity is dubious as well: e.g., if we attempt to measure the quality of our second-order probabilities in the same way, we are threatened with an infinite regress. Perhaps the most serious difficulty, however, is the implausibility of the inferences to which this model gives rise. In brief, the procedure for aggregating probabilistic analyses assumes that they disagree only because of "noise," or random error, in the assessment process; hence, it yields results which do not reflect the possibility that different analyses have drawn on different evidence. We suggest that from a psychological point of view, different analyses may tap different portions of our store of knowledge, even when performed by the same individual. These points are amplified in Cohen et al., 1984, and in a planned paper by Cohen and Lindley.

Summary. Bayesian probability theory is strong in the formal aspects of validity. Its logical foundations are perhaps uniquely compelling in application to the concept of chance. However, the input and computational burdens which it imposes, except when specialized models are adopted, are considerable. It has no adequate resources for representing the quality of an inferential argument, and requires an arbitrary degree of precision in numerical judgments. Even its validity, in a more informal sense, can be questioned. Bayesian theory, as it stands, implies that one's beliefs should be coherent but provides no guidance for choosing among alternative equally coherent analyses. Moreover, by assuming that all assessments are based on the same evidence, it closes off the most promising source of such guidance. We have argued that the application of Bayesian theory to a problem is not linear process in which conclusions are computed from inputs. It is (or often should be) an iterative bootstrapping process in which comparison of conclusions arrived at by different methods leads to revision of inputs and assumptions, until overall plausibility is maximized. This process of revising probability assessments should be guided by a judgment of their quality. A more satisfactory account of completeness of evidence is, therefore, essential.

2.2.2 Belief functions. *Nature of the theory*. In the theory of belief functions introduced by Shafer (1976), Bayesian probabilities are replaced by a concept of evidential support. The contrast, according to Shafer (1981;

Shafer and Tversky, 1983) is between the chance that a hypothesis is true, on the one hand, and the chance that the evidence *means* (or proves) that the hypothesis is true, on the other. Thus, we shift focus from truth of a hypothesis to the evaluation of an evidential argument. As a result, the system (a) is able to provide an explicit measure of quality of evidence, (b) is less prone to require a degree of definiteness in inputs that exceeds the knowledge of the expert, and (c) permits segmentation of reasoning into analyses that depend on independent bodies of evidence.

In Shafer's system, the support for a hypothesis and for its complement need not add to unity. For example, if a witness with poor eyesight reports the presence of an enemy antiaircraft installation at a specific location, there is a certain probability that his eyesight was adequate on the relevant occasion and a certain probability that it was not, hence, that the evidence is *irrelevant*. In the first case, the evidence proves the artillery is there. In neither case could the evidence prove the artillery is not there.

To the extent that the sum of support for a hypothesis and its complement falls short of unity, there is "uncommitted" support, i.e., the argument based on the present evidence is unreliable. Evidential support for a hypothesis is a lower bound on the probability of its being true, since the hypothesis *could* be true even though our evidence fails to demonstrate it. The upper bound is given by supposing that all present evidence that is *consistent* with the truth of the hypothesis were in fact to prove it. The interval between lower and upper bounds, i.e., the range of permissible belief, thus reflects the unreliability of current arguments. This concept is closely related to *completeness* of evidence, since the more unreliable an argument is, the more changeable the resulting beliefs are as new evidence (with associated arguments) are discovered. These concepts are not captured by Bayesian probabilities.

In Shafer's calculus, support $m(\cdot)$ is allocated not to hypotheses, but to <u>sets</u> of hypotheses. Shafer allows us, therefore, to talk of the support we can place in any subset of the set of all hypotheses. In the case of three hypotheses, H_1, H_2 and H_3, for example, we could allocate support to H_1, H_2, H_3, $\{H_1$ or $H_2\}$, $\{H_1$ or $H_3\}$, $\{H_2$ or $H_3\}$, and $\{H_1$ or H_2 or $H_3\}$. As with probability, the total support across these subsets will sum to 1, and each

support m(·) will be between 0 and 1. It is natural, then, to say that m(·) gives the probability that what the evidence *means* is that the truth lies somewhere in the indicated subset.

Suppose, for example, that we know in the case of three hypotheses that H_3 is false, but have no evidence to distinguish between H_1 and H_2. In that case, we would put $m(\{H_1 \text{ or } H_2\}) = 1$, and give zero support to all the other possible subsets. Alternatively, we may feel that the evidence *either* means that H_3 is true, *or* that $\{H_1 \text{ or } H_3\}$ is true, *or* that it is not telling us anything (i.e., $\{H_1 \text{ or } H_2 \text{ or } H_3\}$ is true), and that the weight of evidence is just as strong with each possibility. In that case $m(H_3) = m(\{H_1 \text{ or } H_3\}) = m(\{H_1 \text{ or } H_2 \text{ or } H_3\}) = 1/3$. In a Bayesian analysis, arbitrary decisions would have to be made about allocating probability *within* these subsets, requiring judgments that are unsupported by the evidence.

This same device, of allocating support to subsets of hypotheses, enables us to represent the reliability of probability assessments. Suppose, for example, that the presence of feature X in a signal is associated with an SA-4 70% of the time and with other threats 30% of the time, based on frequency data from a set of previous signal analyses. If we are confident that an image now being analyzed is representative of this set, we may have m(SA-4) = .7 and m(other) = .3. But if there is reason to doubt the relevance of the frequency data to the present problem (e.g., due to possible presence of ECM in the region), we may *discount* this support function by allocating same percentage of support to the universal set. For example, with a discount rate of 30%, we get m(SA-4) = .49, m(other) = .21, and m ({SA-4, other}) = .30. The latter reflects the chance that the frequency data is irrelevant.

Shafer's belief function Bel(·) summarizes the implications of the m(·) for a given subset of hypotheses. Bel(A) is defined as the total support for all subsets of hypotheses contained within A; in other words, Bel(A) is the probability that the evidence *implies* that the truth is in A. The plausibility function Pl(·) is the total support for all subsets which overlap with a given subset.

Thus, Pl(A) equals $1\text{-Bel}(\overline{A})$; i.e., the probability that the evidence does not imply the truth to be in not-A. In one of the examples above, with

$$m(H_3) = m(\{H_1 \text{ or } H_3\}) = m(\{H_1 \text{ or } H_2 \text{ or } H_3\}) = 1/3,$$

we get:

$$Bel(H_3) = m(H_3) = 1/3; \; Pl(H_3) = 1\text{-}Bel(\{H_1 \text{ or } H_2\}) = 1$$

$$Bel(\{H_1 \text{ or } H_3\}) = m(H_3) + m(\{H_1 \text{ or } H_3\}) = 2/3;$$

$$Pl(\{H_1 \text{ or } H_3\}) = 1\text{-}Bel(\{H_2\}) = 1.$$

Dempster's rule. Thus far, we have focused on the representation of uncertainty in Shafer's system. For it to be a useful calculus, we need a procedure for inferring degrees of belief in hypotheses in the light of more than one piece of evidence. This is accomplished in Shafer's theory by Dempster's rule. The essential intuition is simply that the "meaning" of the combination of two pieces of evidence is the intersection, or common element, of the two subsets constituting their separate meanings. For example, if evidence E_1 proves $\{H_1$ or $H_2\}$, and evidence E_2 proves $\{H_2$ or $H_3\}$, then the combination $E_1 + E_2$ proves H_2. Since the two pieces of evidence are assumed to be independent, the probability of any given combination of meanings is the product of their separate probabilities.

Let X be a set of hypotheses $H_1, H_2, \ldots, H_n$, and write 2^X for the power set of X, that is, the set of all subsets of X. Thus, a member of 2^X will be a subset of hypotheses, such as $\{H_2, H_5, H_7\}$, H_3, or $\{H_1, H_2, H_3, H_4\}$, etc. Then if $m_1(A)$ is the support given to A by one piece of evidence, and $m_2(A)$ is the support given by a second piece of evidence, Dempster's rule is that the support that should be given to A by the two pieces of evidence is:

$$m_{12}(A) = \frac{\sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2)}{1 - \sum_{B_1 \cap B_2 = \phi} m_1(B_1) m_2(B_2)}.$$

The numerator here is the sum of the products of support for all pairs of subsets A_1, A_2 whose intersection is precisely A. The denominator is a normalizing factor which ensures that $m_{12}(\cdot)$ sums to 1, by eliminating support for impossible combinations.

Consider, for example, the following two support functions:

Table 2-1

	$m_1(\cdot)$	$m_2(\cdot)$	$m_{12}(\cdot)$
H_1	0.2	0.1	0.344
H_2	0.1	0.3	0.250
H_3	0.3	0	0.172
H_1H_2	0.1	0.3	0.125
H_1H_3	0.2	0	0.063
H_2H_3	0	0.1	0.016
$H_1H_2H_3$	0.1	0.2	0.031

In the third column, we have used Dempster's rule to compute $m_{12}(\cdot)$. For example

$$m_{12}(H_1H_2) = \frac{m_1(H_1H_2)m_2(H_1H_2)+m_1(H_1H_2)m_2(H_1H_2H_3)+m_1(H_1H_2H_3)m_2(H_1H_2)}{1-C}$$

where

$$\begin{aligned} C = {} & m_1(H_1)[m_2(H_2) + m_2(H_3) + m_2(H_2H_3)] + m_1(H_2)[m_2(H_1) + m_2(H_3) + m_2(H_1H_3)] \\ & + m_1(H_3)[m_2(H_1) + m_2(H_2) + m_2(H_1H_2)] + m_1(H_1H_2)m_2(H_3) + m_1(H_1H_3)m_2(H_2) \\ & + m_1(H_2H_3)m_2(H_1) \end{aligned}$$

and so $$m_{12}(H_1H_2) = \frac{0.1x0.3+0.1x0.2+0.1x0.3}{1 - 0.36} = 0.125.$$

Let us now examine the performance, or at least the potential, of Shafer's theory within our evaluation framework.

Feasibility: Quantity of inputs. One of the main difficulties standing in the way of a Bayesian analysis is its complexity. At first sight the Shaferian approach seems simpler, since complicated independence judgments and conditional probability assessments appear not to be required. This appearance is illusory. Support functions must be assessed over not just the hypothesis set, but over the power set of the hypothesis set. With 10

hypotheses, for example, the support distribution has 1,023 elements. For both Bayesian and Shaferian models, the required number of assessments or judgments increases exponentially with the number of events or hypotheses. To see the parallel, compare the Bayesian rule:

$$\Pr[A \text{ or } B] = \Pr[A] + \Pr[B] - \Pr[A]\Pr[B|A]$$

with Shafer's rule:

$$\mathrm{Bel}(\{A \text{ or } B\}) = m(A) + m(B) + m(\{A \text{ or } B\}).$$

In each case, to get an uncertainty measure for a disjunction (i.e., a member of 2^X), we must make one assessment in addition to the measures already assessed for the elements. For Bayesians, the extra assessment is a conditional probability $\Pr[B|A]$; for Shaferians it is the direct evidential support $m(\{A \text{ or } B\})$.

A Shaferian response to this, in parallel with the Bayesian response, is that specialized models may be developed that require far fewer assessments. In fact, the belief function framework admits a variety of interesting special cases: e.g.,

- simple support functions: all support goes either to some one subset or to the universal set X. Either the evidence limits the truth to lie within one particular subset or it is totally unreliable.
- discounted probabilistic support functions: all support goes to individual hypotheses (as in a standard probability distribution), with some additional support possibly going to the universal set X (reflecting a judgment of the quality of the evidence for the probability distribution);
- consonant support functions: all support goes to a nested series of subsets of hypotheses; i.e., the evidence points in a certain direction but is unclear how far we should go;
- hierarchical support functions: the evidence supports subsets of hypotheses that can be arranged in a tree.

Here again, however, (as in the Bayesian case) complex and difficult judgments must be made to determine that a particular specialized model is applicable,

before savings in quantity of assessments can be realized.

The problem for Shaferians may even be deeper. The applicability of Dempster's rule to two bits of evidence E_1 and E_2 is not automatic. It requires rather careful and difficult consideration of a whole set of independence assumptions. We shall return to this point in our discussion of the validity of Shafer's theory.

Computational tractability. Here again the story is parallel to the Bayesian case. The employment of unrestricted belief function models would involve prohibitive computation. As a result, Gordon and Shortliffe (1984) propose to modify Dempster's rule to simplify computation in MYCIN. Shafer (1984a) has argued in response that *ad hoc* modifications of this sort might be avoided by a control strategy that intelligently exploits the structure of restricted belief function models, such as the hierarchical structure proposed for MYCIN. Here as in the Bayesian case, feasibility is purchased only in special cases, and, evidently, at the cost of complex and subtle judgments regarding the structure of the overall argument.

Validity: Semantics. Shafer argues that the requirement for a behavioral specification of probabilities is irrelevant. People bet in a certain way *because* of their beliefs and preferences; observing their own betting behavior will not help them to *assess* those beliefs. Shafer thus urges a shift from the positivist to a more cognitive orientation. He argues that uncertainty is quantified on the basis of an analogy between one's problem and a "canonical example". In Bayesian modeling, we assess the probability of an event by comparing its likelihood with the likelihood of a frequency-based event, such as a random drawing from an urn. Thus, for Shafer, to say that the Bayesian probability of an event is x is to say that it is "like" the chance of drawing a white ball from an urn with a proportion of white balls equal to x. Similarly, to say that your Shaferian belief in a proposition is y, is to compare it to canonical examples in which the reliability of an evidential source is determined by chance.

Unfortunately, Shafer's position is weakened by two considerations: First, his canonical examples, as we shall see below, are far more complex and less obviously usable, even from a cognitive point of view, than the Bayesian

examples. Second, behavioral specification probably plays a *cognitive* role in clarifying the sense of a canonical example. For example, what does it *mean* to say that my uncertainty about whether an object is a building is "like" my uncertainty about drawing from an urn? In what respects must they be similar? Many people will find it illuminating when told it means that I would bet at equal stakes on either event.

A major strength of Shafer's theory, nevertheless, is the naturalness of the input format it imposes:

- Assessments need go no further than the evidence justifies. As we have seen, "ignorance" is naturally represented by assigning support to a subset of hypotheses, with no further commitment to an allocation within the subset. A Bayesian must decide among quite definite and distinct, but equally arbitrary, allocations of probability.
- Weight or completeness of evidence is quite intuitively represented as the degree to which the sum of belief for a hypothesis and its complement falls short of unity.
- Assessments may be based on distinct, separable bodies of evidence, rather than requiring--as in Bayesian theory--that all assessments be based on all the evidence.

Face validity. Belief function theory possesses no deep axiomatic justification comparable to the de Finetti and Lindley arguments for Bayesian theory. Not coincidentally, however, Shafer has offered a view of model "validation" which contrasts sharply with the axiomatic approach. On Shafer's view (1981; Shafer and Tversky, 1983), theories of inference are tools which can be used to help us *construct* (rather than elicit or discover) a set of probabilities. The justification for applying a particular tool to a particular problem is that we see an analogy between that problem and the canonical example underlying the theory. For example, to the extent that the Bayesian theory has anything to contribute, it is by establishing a persuasive analogy between your problem and a situation, like drawing balls from an urn, where the truth is generated by known chances.

Bayesian analogies of this sort, according to Shafer, will usually be imperfect, because in the canonical example we *know* the rules of the game that determine how the truth is generated (e.g., the composition of the urn and the

procedure for drawing a ball). In real problems, there are nearly always many aspects of the situation where comparable rules cannot be given without making numerous assumptions. When these assumptions become very extensive, it may be better to switch to a simpler kind of model, which is more plausible despite not giving a complete picture of how the truth is generated. Such simpler models can be based on canonical examples in which the meaning of the evidence rather than the truth is generated by known chances.

We comment on Shafer's position at two levels: First, how convincing is his *concept* of validity? Second, how plausible or useful are the canonical examples underlying belief functions?

Concept of validity. For Shafer, validity reduces to face validity and plausibility of instances. His argument for this position, however, contains some confusion. Shafer mistakenly assumes that the adoption of an axiomatic framework implies a belief in pre-existing rather than constructed probabilities. Thus, Shafer (1984a) speaks derisively of assessment in the Bayesian context as "pretending" that one already has probabilistically coherent beliefs and preferences, and then, somehow, "trying to figure out what they are."

Our own view is that Shafer is correct to regard probability frameworks as tools for the construction, rather than discovery, of probabilities. But he is wrong in supposing that the axiomatic derivation of a framework detracts from this role--as long as we understand, as argued above, that axiomatic derivation is only *one* argument in favor of a given framework. If taken seriously, Shafer's argument would declare as "non-constructive" any set of prior constraints on the way uncertainty is represented or manipulated; thus, it applies as strongly against belief functions and Dempster's rule as to Bayesian probabilities. The solution in our view is not to drop constraints, but to drop the view that any particular set of constraints is inevitable. Thus, probability assessment as we understand it is an iterative and constructive process, in which a tentative framework (e.g., Bayesian or Shaferian) is adopted, assessments are made within the framework, checked for consistency, and revised; if the overall result is unnatural or implausible, the framework itself may be rejected or revised. In other words, "pretending" that a framework is correct is a legitimate strategy in uncertainty assessment;

indeed, it is the only *possible* strategy. A framework is of use as a tool precisely because it *does* impose (tentative) constraints on the assessments that are produced. It challenges the expert to actively shape a previously disorganized and perhaps even unverbalized set of beliefs. It serves as a medium or language in which the expert "thinks" about uncertainty and in which he expresses those thoughts. A supposedly "neutral" framework, that imposed no format or structure, beyond that already present, would not help the expert in the process of construction and could not advance his or our understanding of his beliefs. (See Cohen, Mavor, and Kidd, 1984, for a more general argument in the context of knowledge engineering.)

In sum, Shafer's argument for a constructive process of probability assessment is correct. But he appears to have drawn two unnecessary conclusions: (1) It in no way contradicts the added plausibility that may be lent to a framework by the existence of an axiomatic derivation; and (2) it should not blind us to the importance of the iterative strategy of tentatively adopting a framework and testing its implications.

Shafer's canonical example. As noted above, when we apply a belief function analysis, we "pretend" that the meaning of the evidence is generated by known chances. In order to evaluate Shafer's theory in terms of face validity, we must examine this analogy more closely. In particular, we must focus on the independence assumptions embodied in the canonical example which are required to license an application of Dempster's rule. It turns out that these assumptions are the primary constraints imposed by Shafer's theory on the process of evaluating evidence; hence, they are its main contribution to the "construction" of probability judgments. They have also been the major source of controversy between Shafer and Bayesians. Early critics of Shafer's work (e.g., Williams, 1978) complained about the obscurity of Shafer's notion of "independent evidence." In a recent paper, however, Shafer (in press) has clarified this concept considerably.

Shafer's interpretation of belief functions involves two sets of hypotheses (or "frames") as shown in Figure 2-1. One frame, S, is a set of background hypotheses which concern the state of the process that produced the evidence at hand. For example, if the evidence E_1 is a witness's testimony that he saw antiaircraft artillery in a certain location, the frame S may simply be the

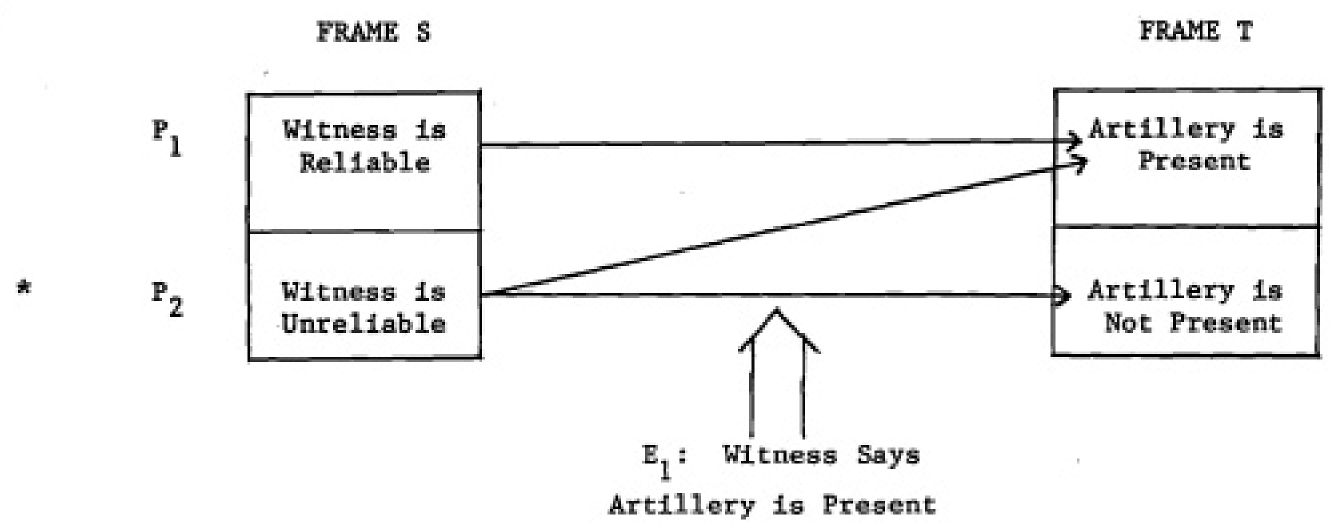


Figure 2-1: Illustration of Canonical Example for Belief Functions

two possibilities (the witness is reliable, the witness is not reliable). The other frame, T, contains the hypotheses of primary interest, e.g., (the artillery is present, the artillery is not present). To get a belief function, we only need (i) a probability distribution over S, i.e., standard probabilities P_1 and P_2, for the reliability and unreliability of the witness; and (ii) a mapping from S to T based on the content of the evidence. Since the evidence is the witness's report of artillery, reliability in S maps onto (the artillery is present) in T; unreliability in S maps onto the set (the artillery is present, the artillery is not present) in T. Support m(A) for a subset A in T is just the probability for hypotheses in S that map only onto A. (We have referred to this, somewhat loosely, as the probability that the evidence "means" A). Bel(A) for a subset A in T is the sum of the probabilities for hypotheses in S that map onto subsets of T that are contained in A. Thus, in our example, Bel(artillery is present) = P_1; Bel({present, not present}) = $P_1 + P_2$.

Suppose we now receive a second piece of evidence, E_2, which is the testimony of a second witness that he saw artillery in the same vicinity. We define a new belief function for this witness by specifying a frame S_2 with the elements (the second witness is reliable, the second witness is unreliable), and by assessing probabilities P_1' and P_2' over S_2. What is our new overall belief in the elements of T? Naming S as S_1, Figure 2-2 shows a new frame, $S_1 x S_2$, which results from combining elements of S_1 and S_2. Each cell has a probability which is the product of the probabilities of the elements from S_1 and S_2; and each cell is mapped onto a subset of hypotheses in T, based on knowledge of E_1 and E_2. According to this mapping (as shown by the labels in the cells), support for the artillery being present equals the chance that either witness 1 or witness 2 is reliable, i.e., $P_1P_1' + P_1P_2' + P_2P_1'$. This is the result given by Dempster's rule.

What if the report of the second witness contradicts, rather than confirms, the first? That is, E_2 is a report that artillery is not present in the specified location. In that case, the new frame, $S_1 x S_2$, appears as in Figure 2-3. The only change is in the mapping of the cells to subsets in T--a change required by the change in E_2. It turns out, however, that the cell corresponding to both witnesses being reliable does not map to any subset in T. Since E_1 and E_2 are contradictory, both cannot be true. Thus, we use our

FRAME S_1 X S_2

* Witness 2		Witness 1: Reliable (P_1)	Witness 1: Not Reliable (P_2)
	Reliable (P_1')	Artillery Present (P_1P_1')	Artillery Present (P_2P_1')
	Not Reliable (P_2')	Artillery Present (P_1P_2')	{Artillery Present, Artillery Not Present} (P_2P_2')

Figure 2-8: Canonical Example for Combination of Concurring Evidence

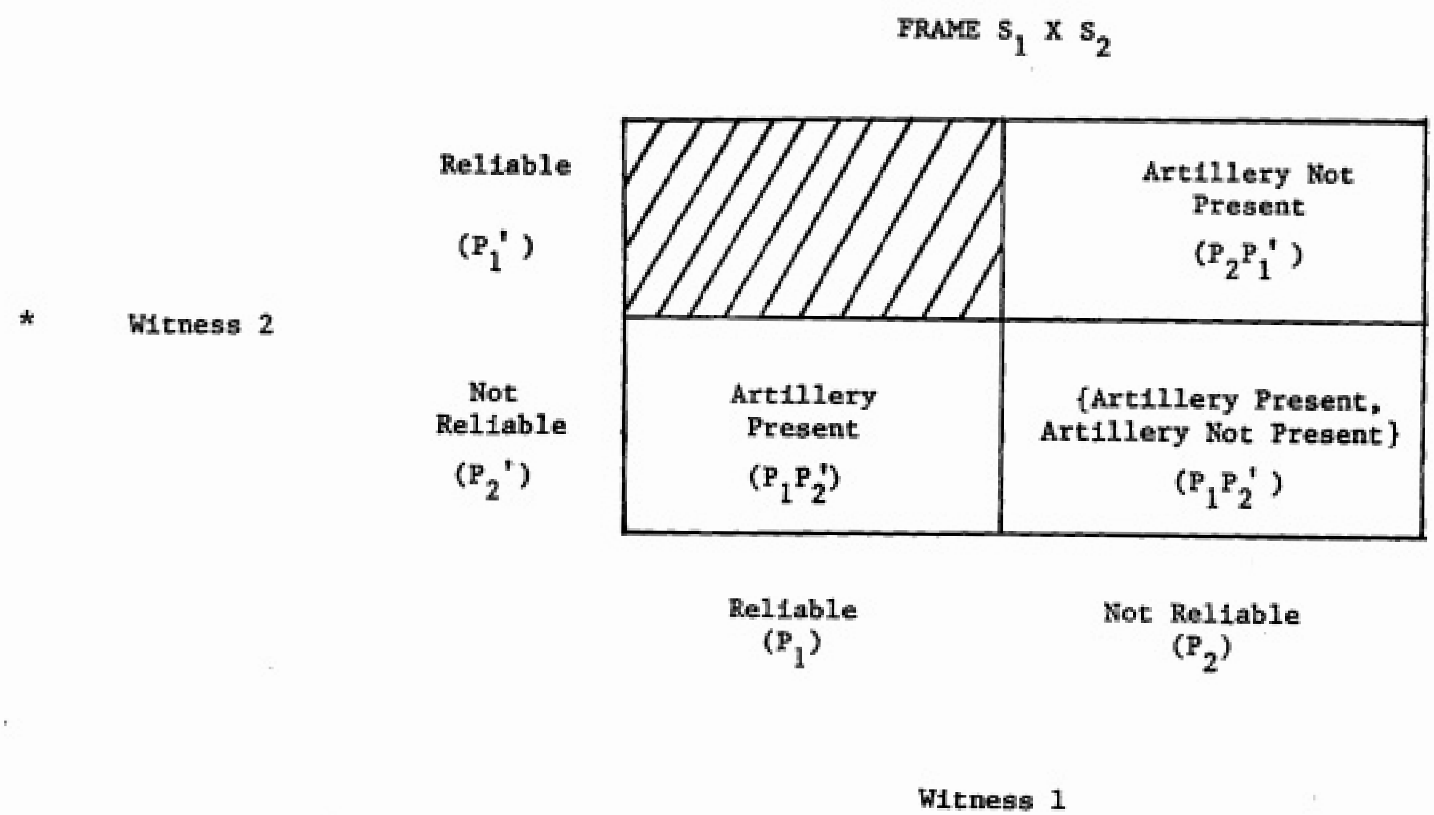


Figure 2-3: Canonical Example for Combination of Conflicting Evidence

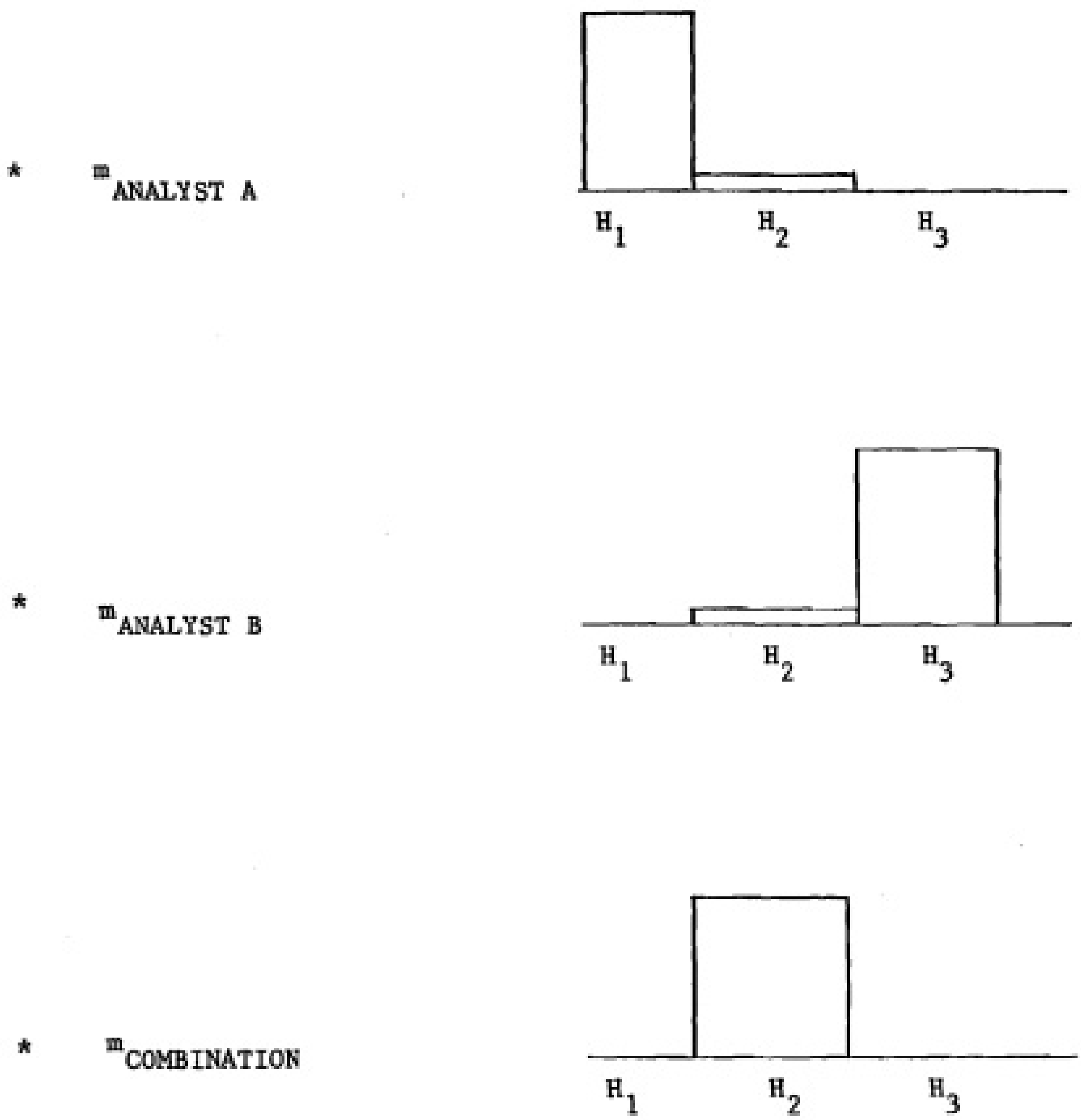


Figure 2-4: Support Functions to Illustrate Combination of Conflicting Evidence by Dempster's Rule

knowledge of E_1 and E_2 to prune out impossible cells in $S_1 x S_2$. According to the mapping, support for artillery being present equals the chance that witness 1 is reliable and witness 2 is unreliable, i.e., $P_1 P_2'/(1-P_1 P_1')$, normalizing to remove the impossible case. Once again, this is the result of applying Dempster's rule.

In many of Shafer's discussions, he appears to argue that Dempster's rule is justified in situations which "resemble" this canonical example, because it is the correct rule for the example (just as Bayesian rules are correct for the case of drawing balls from an urn). But what makes it correct? Even these simple examples may seem too complex for such a direct appeal to intuition. A recent paper by Shafer (in press) contains a more extensive discussion of the preconditions of Dempster's rule. We can use Dempster's rule, he says, only if the following judgments are made:

(a) Before consideration of the mapping to T, any hypothesis in S_1 is compatible with any hypothesis in S_2 (so $S_1 x S_2$ can be defined as a new frame).

(b) Probabilities for elements of S_1 are independent of elements in S_2 (e.g., we do not alter our estimate of the reliability of one witness based on the reliability or unreliability of the other witness).

(c) If we could draw a conclusion about the truth of a subset in T by knowing that a certain combination of hypotheses from S_1 and S_2 was the case, then we could have drawn the same conclusion by knowing that either one or the other of the hypotheses (from S_1 or S_2) was the case. (In the example of concurring witnesses, we can conclude that artillery is present if both witnesses are reliable; but all we needed was one or the other to be reliable).

(d) The evidence we use for assessing S_1 and S_2 tells us nothing more directly about T. (All the work of reasoning about T is transferred to reasoning about S.)

Having enumerated these assumptions, we must remark that our original question about the rationale for Dempster's rule remains unanswered. It has not been demonstrated in any way that Dempster's rule "follows from" these preconditions. Perhaps Shafer means simply that when these particular conditions are met, Dempster's rule will appear more plausible or natural.

Note, however, that the canonical situation described by these conditions in-

cludes a chance model: Because of assumptions (a) and (b), the probability for a component of $S_1 x S_2$ is simply the product of the probabilities assigned to the components of S_1 and S_2. It is tempting, therefore, to view the belief function model as a special case of a Bayesian analysis, defined by the restrictions outlined in (a) - (d). In that case, Dempster's rule should be justifiable from (a) - (d) by the rules of probability theory. Moreover, Shafer's model would then inherit the axiomatic justification of the Bayesian model in the special circumstances where it applied.

A Bayesian foundation for belief functions? To see how this might work, consider the simple case of Figure 2-2, with H = the artillery is present, $\overline{H}$ = the artillery is not present, R = the first witness is reliable, and $\overline{R}$ = the first witness is not reliable. It follows from probability theory that:

$$Pr(H) = Pr(H|R)Pr(R) + Pr(H|\overline{R})Pr(\overline{R}).$$

Following Shafer's definitions, we interpret m(H) as Pr(R) and m(H or $\overline{H}$) as $Pr(\overline{R})$. In addition, from our knowledge of E_1 (i.e., the mapping from S_1 to T which it establishes), and using (d), we know that Pr(H|R) = 1; if the witness is reliable, then the artillery is present. Hence, we may write

$$Pr(H) = m(H) + Pr(H|\overline{R})\ m(H \text{ or } \overline{H})$$

and this gives

$$Bel(H) = m(H) \leq Pr(H) \leq m(H)+m(H \text{ or } \overline{H}) = Pl(H),$$

where Bel(H) and Pl(H) are Shafer's belief and plausibility functions. It *appears*, then, that the belief function analysis is simply an incomplete Bayesian analysis. Our uncertainty about Pr(H) is due to our failure, in the belief function approach, to specify $Pr(H|\overline{R})$, i.e., the chance of the hypothesis being true *despite* the fact that the present evidence is unreliable. This is just another way of saying that Shafer is interested in the proof of the hypothesis, not its truth. If $Pr(H|\overline{R}) = 0$, Pr(H) = Bel(H); and if $Pr(H|\overline{R}) = 1$, Pr(H) = Pl(H). Thus, Bel(H) and Pl(H) give lower and upper bounds for the Bayesian probability.

Let us now see how Dempster's rule works within this Bayesian interpretation. Let R_1 and R_2 refer to the reliability of the first and second witness, respectively, and take the case where E_1 and E_2 agree. A Bayesian probability $Pr(\cdot|\cdot)$, is a function of <u>two</u> arguments, the event and the evidence. Presumably, therefore, in using Dempster's rule, the probability to be bounded is $Pr(H|E_1,E_2)$. Let us for the moment, however, ignore this consideration and use $Pr(H)$. (Note that in the case of one piece of evidence, we likewise used $Pr(H)$ instead of $Pr(H|E_1)$.) By probability theory, we have

$$Pr(H) = Pr(H|R_1 \text{ or } R_2)Pr(R_1 \text{ or } R_2) + Pr(H|\overline{R_1 \text{ or } R_2})Pr(\overline{R_1 \text{ or } R_2}).$$

Substituting based on conditions (a) and (b), we have

$$Pr(H) = Pr(H|R_1 \text{ or } R_2)[Pr(R_1)+Pr(R_2)-Pr(R_1)Pr(R_2)] + Pr(H|\bar{R}_1\bar{R}_2)Pr(\bar{R}_1)Pr(\bar{R}_2).$$

By Dempster's rule,

$$m_{12}(H) = Pr(R_1) + Pr(R_2) - Pr(R_1)Pr(R_2)$$

and by Shafer's definitions,

$$m_{12}(H \text{ or } \bar{H}) = Pr(\bar{R}_1)Pr(\bar{R}_2).$$

Using (c) and (d) and the mapping from S_1xS_2 to T, $Pr(H|R_1 \text{ or } R_2) = 1$. Therefore,

$$Pr(H) = m_{12}(H) + Pr(H|\bar{R}_1\bar{R}_2)m_{12}(H \text{ or } \bar{H}).$$

It follows that

$$Bel_{12}(H) = m_{12}(H) \leq Pr(H) \leq m_{12}(H) + m_{12}(H \text{ or } \bar{H}) = Pl_{12}(H).$$

Thus, Bel(H) and Pl(H), when computed by Dempster's rule, continue to give upper and lower bounds for Pr(H). (Note, however, that $Bel(\cdot)$ and $Pl(\cdot)$ are not bounds on what the future probability *could* be, given further evidence. They are bounds on $Pr(\cdot)$ implied by our *present* evidence.) A similar demonstration can be given for the case where E_1 and E_2 conflict. This approach can be

generalized to the case where support is assigned to arbitrary subsets of hypotheses by regarding "reliability" as a set of separately assessed skills involved in discriminating subsets of hypotheses from their complements.

The problem, of course, is that we have not justified Dempster's rule as a bound on the *Bayesian* probability, $Pr(H|E_1E_2)$. When we conditionalize on the evidence, as we certainly must in a Bayesian analysis, $Pr(R_1 \text{ or } R_2)$ is replaced by

$$Pr(R_1 \text{ or } R_2|E_1E_2) = Pr(R_1|E_1E_2) + Pr(R_2|E_1E_2) - Pr(R_1|E_1E_2)Pr(R_2|E_1E_2R_1).$$

This brings out a curious and critical feature of Shafer's theory. He is asking us to assess the reliability of a witness (or more generally, the status of an evidentiary process) without taking into account our knowledge of what the witness said. In Shafer's canonical example, knowledge of the evidence enters in only for the mapping from S to T, *after* all the probability work has been done on S. In a Bayesian analysis, on the other hand, the credibility of a witness can be shown to depend both on what is said and on its prior probability, i.e., our original tendency to think it true. If a witness says something which is independently believable, our estimate of his reliability increases. More importantly, perhaps, the credibility of one witness can, in a Bayesian analysis, be increased by corroboration of a second witness, and decreased by contradiction.

Assumption (b) is plausible only in light of this restriction. The strict Bayesian version of (b) is

$$Pr(R_2|E_1E_2R_1) = Pr(R_2|E_1E_2).$$

Note that E_1R_1 implies H, i.e., if witness 1 is reliable and says H, H is true. But we would expect, quite generally, that $Pr(R_2|E_2H) > Pr(R_2|E_1E_2)$, i.e., learning for a fact that what the witness said is true increases his credibility more than corroboration by a second witness. On the other hand, if we are assessing a witness's reliability prior to (or without consideration of) his testimony, it does make sense to require that his reliability be independent of the reliability of another witness. We thereby preclude shared

uncertainties (e.g., a conspiracy) in the two evidential processes being combined.

A group of Swedish researchers, whose work is summarized and extended in Freeling and Sahlin (1983), and Freeling (1983), has explored issues such as this. Like Shafer, they focus on the reliability of the evidence, rather than the truth of the hypothesis, i.e., they reject the traditional Bayesian effort to model the chance of a hypothesis when the evidence is unreliable. But unlike Shafer, they analyze reliability in the light of the evidence, as $Pr(R|E)$ rather than $Pr(R)$. In effect, this is an effort to give a proper Bayesian account of the notion of quality or completeness of evidence, rather than truth. (As such, it is an alternative to the idea of second-order probabilities discussed above) The upshot of this research is that if $m(H)$ is equated with $Pr(R|E)$, Dempster's rule cannot in general be justified. Depending on the character of the belief functions being combined, and the kinds of conditional dependence assumed in the Bayesian analysis, Dempster's rule may be correct, a good approximation, or entirely off the mark in comparison to the "proper" Bayesian rule of combination.

While it fails to fully validate Dempster's rule, the Swedish work also lacks most, if not all, of the virtues of the belief function representation. In terms of feasibility, formulations which conditionalize on the evidence become extremely complex even for the simplest examples. The Swedish group has made little progress in deriving rules for the combination of evidence involving the full range of cases to which Dempster's rule applies, in particular, where varying degrees of support are assigned to arbitrary subsets of hypotheses. Moreover, the requirement to assess prior probabilities is incompatible with the segmentation of evidence which is vital for the naturalness of inputs in Shafer's system.

Shafer (in press) explicitly rejects the attempt to provide any sort of Bayesian foundation for belief functions. Arguments based on Dempster's rule "have their own logic"--based on the appropriate canonical examples and an intuitive conviction that the appropriate conditions of independence are satisfied. As noted above, Shafer's appeal to intuition has not entirely succeeded in making that "logic" clear. We propose, however, that it can be clarified. In opposition to both Shafer and the Bayesians, we would argue the

merits of the pseudo-Bayesian analysis of Bel(·) and Pl(·) as bounds on Pr(·), which we illustrated in this section. It fails to derive Dempster's rule as a *special case* of probability theory. Nonetheless, it clarifies the relationship of Dempster's rule to the canonical example, by an argument that *resembles* a valid Bayesian argument in most respects. Moreover, the dissimilarity can be crisply and clearly stated: the argument concerning reliability is conducted without consideration of the content of the evidence. The latter can be regarded as an explicit decision, justified by enormous gains in the simplicity and power of the calculus. This is not equivalent, however, to a fixed *belief* that the content of evidence is irrelevant. In an iterative, bootstrapping system, we can guard against the pitfalls of that assumption by continually reexamining it as an analysis proceeds. In Section 3 we explore the design of a system in which the function of recalibrating sources of evidence in light of corroboration or conflict is assigned to a process of qualitative reasoning.

Role of the assumptions in constructing an analysis. Conditions (b) and (c) play an important role as constraints in the construction of a belief function analysis. Violation requires reassessment of the overall structure of an analysis, redefining frames for either S or T or both (cf., Shafer, 1984a). (c) says that elements from both witnesses' testimony must not be required in order to construct a chain of reasoning that gets us to T. For example, if one witness said p and the other said $p \rightarrow q$, we would need to assume *both* were reliable to infer q. Therefore, these two statements must be counted as parts of a single evidential argument. In this sense, Dempster's rule combines self-contained "arguments" rather than "bits" of evidence. And application of the rule presupposes a more global process of reasoning addressed to problem structuring.

(b) and (c) represent a limitation on Dempster's rule in a second sense: Once our evidence has been segmented into independent arguments, we can combine it by Dempster's rule, but that rule tells us nothing about how two dependent pieces of evidence should be combined *within* a self-contained argument. Clearly, in any expert system application, Dempster's rule must be supplemented by other forms of inference. Interestingly, in a recent paper, Shafer (1984) himself suggested that expert systems will have to make provision for dependent evidence, and that the full range of Bayesian operations can be ap-

plied on probabilities for the background frame, S. This is a departure from the position that only Dempster's rule is appropriate for combining evidence in the belief function context.

We have now noted three different ways in which an expert system application of Shafer's system might need to be supplemented:

- recalibration of sources of evidence in terms of the content of the evidence,
- reframing evidence and hypotheses to achieve independence of arguments, and
- reasoning about dependent evidence within an argument.

We may refer to this set of issues as the *incompleteness* of Dempster's rule, in analogy to the incompleteness of Bayesian theory discussed above.

Plausibility of instances: Conflict of evidence. To what extent does belief function theory yield inferences which are intuitive and plausible in specific applications? A topic of special concern in this regard is conflict of evidence. Zadeh (1984) recently raised an example of the following sort. Suppose we have two experts who we believe to be very reliable and who produce conflicting judgments. For example, there are three possible interpretations of an object x in a specified location: H_1--x is an SA-4 installation; H_2--x is an SA-7 installation; H_3--x is not a threat. Analyst A, using photographic evidence, assigns .99 support to H_1 and .01 to H_2; analyst B, using independent intelligence information, assigns .99 support to H_3 and .01 to H_2. We have the following two support functions, and may combine them by Dempster's rule, as shown in Figure 2-4:

Table 2-2

	$m_A(\cdot)$	$m_B(\cdot)$	$m_{AB}(\cdot)$
H_1	0.99	0	0
H_2	0.01	0.01	1.00
H_3	0	0.99	0

The counterintuitive result, according to Zadeh, is that exclusive support is now assigned to H_2, a hypothesis that neither expert regarded as likely. Moreover, the result is independent of the probabilities assigned to H_1 or H_3.

Shafer's response (in press) is cogent, but ultimately, we feel, off the mark. If we really regard these experts as perfectly reliable, Shafer says, the argument as stated is correct. After all, A says that H_3 is impossible, and B rules out H_1; that leaves H_2 as the only remaining possibility. (It is important to note that exactly the same result would be obtained in Bayesian updating, if we interpret the $m(\cdot)$ as likelihoods of the evidence given the hypothesis and assume that prior probabilities for the three hypotheses are equal.) On the other hand, Shafer argues that experts are seldom in fact perfectly reliable. A more reasonable procedure would be to "discount" the belief functions supplied by the experts to reflect our degree of doubt in the reliability of their reports. In discounting, we reduce each degree of support by a fixed percentage, and allocate the remainder to the universal set $\{H_1,H_2,H_3\}$. The result of applying Dempster's rule will now be a belief function that assigns support to all three hypotheses.

Let us examine this response in a bit more detail. Recalling that we regard these experts as highly reliable (though not perfect), suppose we discount A's belief function by 1% and B's by 2%. The result is the following, as depicted in Figure 2-5:

Table 2-3

	$m_A(\cdot)$	$m_B(\cdot)$	$m_{AB}(\cdot)$
H_1	0.9801	0	.656
H_2	0.0099	0.0098	.013
H_3	0	0.9702	.325
$\{H_1,H_2,H_3\}$	0.01	0.02	.007

We now have a "bimodal" belief function, with the preponderance of support going to H_1 and H_3. This appears, at first look, to be an intuitively plausible result: it reflects our feeling, which we represented in the form

* $m_{ANALYST\ A}$

* $m_{ANALYST\ B}$

* $m_{COMBINATION}$

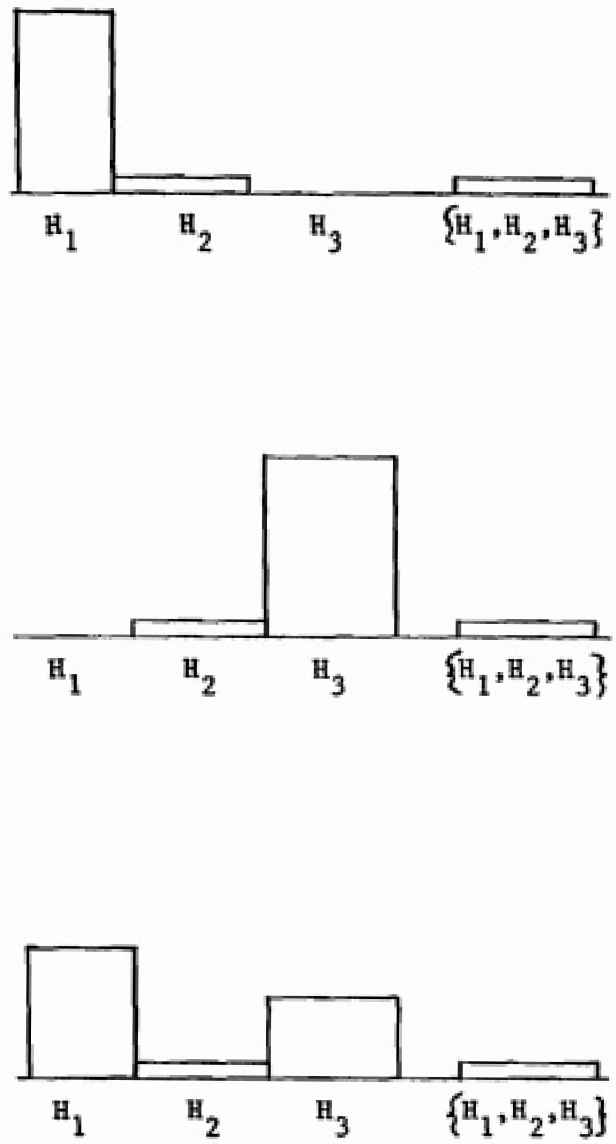


Figure 2-5: Support Functions to Illustrate Combination of Conflicting Evidence with Discounting

of discount rates, that A *or* B (or both) could possibly be unreliable. But let us look a little more closely.

The first thing to note is what a vast difference a small amount of discounting makes. In Table 2-2, after combination by Dempster's rule, there was exclusive support for H_2. In Table 2-3, final support for H_2 is only slightly greater than 1%. The second thing to notice is the large discrepancy between $m_{AB}(H_1)$ and $m_{AB}(H_2)$. Although we did in fact discount B at twice the rate as A, the actual numbers (2% and 1%, respectively) and the difference between them was very small. It is by no means clear that the resulting difference in support for H_1 and H_3 is intuitively plausible. More to the point, the sensitivity of the result for all three hypotheses to very small differences in discount rates is disturbing. Finally, to dramatize the sensitivity even further, note that if support for (H_1,H_2,H_3) were 0 for both experts, and if A assigned 0 support to H_3, and B assigned 0 support to H_1, these very small changes render Dempster's rule indeterminate.

Perhaps the problem is that our original assessment of the reliability of the experts was mistaken. Suppose then we discount A by 29% and B by 30%. We now get:

Table 2-4

	$m_A(\cdot)$	$m_B(\cdot)$	$m_{AB}(\cdot)$
H_1	.7029	0	.4243
H_2	.0071	.007	.0085
H_3	0	.693	.4044
(H_1,H_2,H_3)	.29	.30	.1751

Support for H_1 and H_2 after combination is now roughly equal, certainly a more intuitive result. Then should we have discounted A and B more in the first place? According to Shafer, presumably, this is indeed the case; the fault is not in the theory, but in the initial allocation of support. The example, however, highlights a deeper problem. As we noted above, reliability is to be assessed *as if* we had no knowledge of the evidence actually provided. Thus,

we are apparently not permitted to *use* the conflict between A and B as a clue regarding their capabilities or as a guide to the appropriate amount of discounting. We return to this issue very shortly.

Zadeh himself objects to the procedure in Dempster's rule of normalizing support measures to eliminate impossible combinations. But we think this objection is mistaken. Normalization is in fact the *only* way in Shafer's theory (albeit quite indirect) that our knowledge of the evidence enters into the assessment of reliability. It accomplishes a sort of *de facto* discounting as a function of conflict of evidence. Note in the earlier example of Figure 2-3 that the reliability of witness 1, after combining his testimony with the conflicting evidence of witness 2, is $P_1P_2'/(1-P_1P_1')$. This is less than P_1, the original assessment of witness 1's reliability.

Although normalization is in itself not problematic, nevertheless, it is not a complete or adequate solution to the problem of conflict. First, because there is no lasting effect on later problems, i.e., we have not truly updated our estimate, P_1, of A's reliability in the light of his conflict with B. Second, there is no procedure for exploring potential *reasons* for the conflict. A closer examination of (a) the factors that determined our original reliability estimates, (b) our assumptions regarding independence of the two arguments, and (c) the internal structure of the arguments employed by A and B, might lead to a revision in beliefs and assumptions that permanently improves our knowledge base.

We argue, then, that the revision of reliability estimates is only one possible result of an iterative, constructive process of problem solving prompted by conflict of evidence. (We also have the options of reframing evidence and hypotheses to reflect revised judgments of independence and of revising specific beliefs internal to the conflicting arguments. Therefore, such revisions must be justified by considerations which, once discovered, carry weight independent of the conflict of evidence that led to their discovery. Ideally, these newly discovered factors could be regarded as sufficient to justify revisions in reliability estimates independently of E_1 and E_2. (Referring to these factors as F, we would have $Pr(R_1|E_1E_2F) = Pr(R_1|F)$.) This justifies the reassessment of reliabilities in the light of the evidence in the Shafer-

Dempster system, and is the method used in the inference framework to be described in Section 3.

What is "conflict of evidence"? So far, we have taken for granted the notion of conflicting evidence, and that in some cases at least special steps are justified in dealing with it. But it is by no means obvious what "conflict" is, or why steps outside the normal calculus of uncertainty should be required to handle it. Conflict of evidence does not appear, on the surface, to be the same as incoherence. The formal constraints of Bayesian theory dictate, as we saw above, that multiple probabilistic analyses should agree with one another and with direct judgment. Similar coherence constraints can be derived for Shafer's theory from the requirement that uncertainty on S be measured by a probability. But it is implicit that these analyses are, or should be, based on the same evidence. There appears to be no corresponding guarantee or prescription that arguments based on different evidence should arrive at the same or similar conclusions. Dempster's rule is designed explicitly to combine arguments based on independent evidence; hence, there are no direct constraints on the extent to which those arguments must agree (except that there be at least one pair of meanings from the two arguments whose intersection is non-empty).

Nevertheless, we propose that the resolution of conflict in a belief function analysis be construed as a desire for coherence. The missing element, which is responsible for the incoherence, is a judgment, often implicit, regarding the overall structure which the final belief representation is expected to have. Such judgments are based on one's knowledge about reasoning in a particular problem domain. "Conflicting evidence" is evidence whose combination produces a structure that violates such a prior expectation. Thus, the definition of "conflict" will vary from one problem domain to another. The locus of conflict is not, strictly speaking, between the two sources of evidence, but between both of them, on one side, and a structural expectation regarding the outcome of the argument, on the other. When a conflict of this sort occurs, in an iterative, constructive context, the decision maker has a choice of either revising the expectation or else making one or more of the three kinds of changes we discussed above (revising discount rates, frames, or steps in an argument).

If belief functions are probabilistic with discounting (i.e., assign support only to single hypotheses and to the universal set), then it is often plausible to require that hypotheses which receive very little support from either of two arguments not receive predominant support in the combined analysis. This was the basis of the adjustment of discount rates in the above example (and also seems to underlie the use of discounting in Shafer, 1982). Note that an analogous requirement is recommended for Bayesian analysis by DeGroot (1982).

Other possible structural expectations regarding the form of a belief function model include that it be consonant or hierarchical. In these cases, support is assigned only to nested subsets of hypotheses or to subsets that form a tree, respectively. Neither of these properties is necessarily preserved through combination by Dempster's rule. Yet, as we noted above, such structural constraints may (a) be quite plausible for particular problem domains (cf., Gordon and Shortliffe, 1984, on medical diagnosis), and (b) be required to improve the computational tractability of a Dempster-Shafer model. Thus, once again, a higher-order process of qualitative reasoning may be necessary to explore revisions in beliefs and assumptions, in order to handle "conflict" and to ensure the applicability and plausibility of a Dempster-Shafer calculus (see Section 3 below).

An important by-product of requiring consonance should be noted. One potential criticism of Shafer's theory is that it lacks a concept of the *acceptance* of a hypothesis once it achieves a sufficient degree of evidential support (e.g., Levi, 1983; L.J. Cohen, 1977). A precondition of acceptance--and what makes it a useful concept in some contexts--is that it should yield a logically consistent and complete story. Neither is true if a threshold or cutoff for acceptance is defined on Bel(·) in Shafer's system. Both a hypothesis and its complement could have positive support, and thus conceivably both could be accepted, yielding a contradiction. Moreover, two propositions, p and q, might be accepted but their conjunction, p&q, rejected. Both of these problems disappear in a consonant belief function: Since a hypothesis and its complement are not nested, they cannot both receive support; and it can be shown that Bel(p&q) = MIN(Bel(p),Bel(q)) and thus that a conjunction is at least as credible as either of its conjuncts.

In all these cases, there is a tension between the desirability or plausibility of depicting the state of evidence "as it is," conflicts and all, and attempting to produce a resolution or reconciliation within the framework of some plausible or desirable global requirement. We claim that this tension is at the heart of any truly intelligent and flexible reasoning with probabilistic systems.

Summary. Shafer's theory provides a natural representation of quality of evidence and relaxes the assessment requirement to the extent that the evidence is incomplete. Like Bayesian theory, however, belief function models impose inordinate input and computational demands unless specialized models are adopted. The validity of Shaferian theory has not been clearly established, although it may be illuminated by a partial Bayesian derivation. A major difference is that Shafer's theory does not permit reassessment of the quality of an information source in terms of what that source says; the credibility of one witness cannot be increased by corroboration of a second witness or decreased by contradiction. In belief function theory, the outcome of combining the information from two conflicting data sources can vary dramatically, depending on our assessment of their credibility. Yet we cannot use the two sources to crosscheck one another. We argue that this gap in Shafer's theory requires that it be supplemented by a process of qualitative reasoning that reexamines sources of evidence as an analysis proceeds, and recalibrates them in the light of corroboration or conflict. The same process might supplement Shafer's theory in other ways: by reframing evidence and hypotheses to establish independence of evidential arguments, and by revising inferential steps which are internal to such arguments.

2.2.3 Fuzzy set theory. *Nature of the theory*. Since L.A. Zadeh advanced fuzzy set theory in 1965, an enormous amount of interest, and a very large literature, has been generated. Most of this interest has been theoretical, concerned with the mathematical implications of the theory, but there have been a number of attempts to apply the theory to practical problems. This is in line with Zadeh's original reason for introducing the concept. He argued that much systems analysis was inadequate because its requirements were too precise. He felt that our intuitive understanding of concepts and, more interestingly, our reasoning about those concepts, were typically imprecise, yet analysis (especially with computers) required precisification. To resolve

this paradox, he introduced the now well-known concept of the fuzzy set--a set with imprecise boundaries. The essential element is the membership function $\mu_A(x)$ which represents the degree to which an element x belongs to some set A. If $\mu_A(x) = 1$ then x indisputably belongs to A, while if $\mu_A(x) = 0$, x does not belong to A. An intermediate value, such as $\mu_A(x) = 0.6$, indicates that x belongs to the set to some degree. Fuzzy sets are thus a precise tool for representing and manipulating imprecise notions.

Application of fuzzy set theory involves: first, the representation of imprecise concept by fuzzy sets; second, the use of a calculus to construct other fuzzy sets representing the output variables in an analysis; and third, reinterpretation of the results in imprecise language (see L.A. Zadeh, 1975). The first and last steps are crucial if the flavor of the fuzzy theory is to be fully captured. The core idea is to construct a calculus for the formal (i.e., *precise* manipulation of imprecise concepts, which takes in imprecise inputs and puts out imprecise outputs.

Applications of fuzzy set theory to inference. The theory of fuzzy sets can be applied in many ways, in the sense that wherever a mathematical relationship exists, it can be fuzzified. Thus, there are many possibilities for using the fuzzy calculus in conjunction with other inference theories. Alternatively, it can be applied directly to ordinary imprecise reasoning (by experts or non-experts) in natural language. We will introduce some of the formation of fuzzy set theory by examples of these two types.

Fuzzy Logic. In fuzzy set theory, the statement,

"The installation is large,"

could be represented as a fuzzy membership $\mu_L(i)$, which measures the degree of membership of the installation i in the set of "large" installations (where 0 represents non-membership and 1 denotes complete membership). The degree to which an installation is both large and modern is the minimum of the two membership functions:

$$\mu_{LM}(i) = \min(\mu_L(i), \mu_M(i)).$$

Implication in fuzzy set theory is defined as a relation. Thus, "if U is F, then V is G," where F and G are fuzzy sets on the variables u and v underlying U and V, is described by the relation

$$\mu_{V/U}(u,v) = \min(1,\mu_2(v) + 1- \mu_1(u))$$

using an obvious notation. This may be interpreted as the extent to which a particular value of U implies a particular value of V.

The next step is to combine the rule with a statement about the fact described in its antecedent. In fuzzy implication, not only may be the concepts involved be fuzzy, but the match between a fact and the antecedent of a rule may be a matter of degree as well. Thus, we may have a rule stating "If U is F then V is G," but an input stating that "U is F*". where F and F* are not the same. Zadeh defines this as

$$\mu_Y(v) = \max_u(\min(\mu_{F*}(u), \mu_{V/U}(u,v))).$$

where Y is the fuzzy set that results from combining F* and V/U.

Moving back to our example, suppose we have a rule,

"If an installation is modern, then the danger is high."

We could express this rule as

$$\mu_{I/D}(i,d) = \min(1, \mu_H(d) + 1 - \mu_M(i)).$$

the extent to which modernity of an installation implies high danger of the installation. Now, suppose we have another fuzzy membership function $_R(i)$ representing, perhaps, the input, "the installation was built recently." The result is

$$\begin{aligned}\mu_y(d) &= \max_i(\min(\mu_R(i), \mu_{I/D}(I,D))) \\ &= \max_i(\min(\mu_R(i), \min(1,\mu_H(d) + 1 - \mu_M(i)))).\end{aligned}$$

This output can be interpreted as a quantitative measure that the danger is high, given the fuzzy evidence regarding modernity and the fuzzy implication rule. The output may now be translated into an imprecise natural range representation (e.g., "danger is quite possibly high").

Fuzzy probabilities. Uncertainty about facts (i.e., chance) was not mentioned above; we just talked about imprecision. Zadeh stresses that the two concepts are distinct, and that fuzzy set theory should only be used to describe imprecision. If we are imprecise our uncertainties, however, then a role exists for describing that imprecision with fuzzy sets. Watson et al. (1979) and Zadeh (1981) discuss this idea in the context of decision analysis, but it can clearly be applied to any use of Bayesian probability theory, or belief function theory.

The basic tool for fuzzifying a calculus is Zadeh's extension principle, which enables us to compute the fuzzy set membership function for a variable when it is a function of variables whose fuzzy set membership functions are known. Let $Y = F(X_1,X_2,\ldots,X_n)$. Then $\mu_Y(y) = \max[\min(\mu_{x_1}(x_1), \mu_{x_2}(x_2),\ldots, \mu_{x_3}(x_3))$ where $_Y(y)$ is the extent to which a value y belongs to the set of possible numbers for the output variable.

Suppose a threat classification procedure leads to a probability p that a threat should be classified as an SA-4. Imagine we have a loss function which gives unit loss if misclassification occurs, and zero loss if not. Then the expected loss from classifying the object as an SA-4 is

$$1 \times (1-p) + 0 \times p = 1-p$$

while the expected loss from classifying the object as 'not an SA-4' is

$$1 \times p + 0 \times (1-p) = p.$$

Clearly, we minimize expected loss by categorizing it as an SA-4 if $p>1/2$. Now suppose that we are imprecise about p to the extent that we can only describe a fuzzy set $\mu(p)$ about possible values of p. Fuzzy sets for the expected loss in the two cases (actually $\mu(1-p)$ and $\mu(p)$) can be produced using Zadeh's extension principle. But what conclusions can we draw? Freeling

(1980) discusses this in some detail, suggesting several alternative approaches. As we might expect, when results are fuzzy, the analysis may not indicate any particular decision regarding classification.

As with the Bayesian analysis, there are some non-trivial problems in attempting to apply fuzzy set theory to inference in expert systems.

Feasibility. We criticized both Bayesian theory and belief function theory on the grounds that the analysis involved in practical problems can be quite complex. This will also be true of fuzzy set theory. The fact that functions of variables have to be handled in computations makes the analysis difficult to handle numerically. Nonetheless, there are indications that the max-min operations are numerically easier than the sum-product operations of the other theories. It would be wrong, however, to assert that the use of fuzzy set theory removes all of the difficulties caused by complexity in the other two theories examined here.

Validity. For a theory which has had an enormous literature, there is still a considerable discussion amongst scholars on the justification and interpretation of the theory.

Semantics: <u>Where do the numbers come from</u>? This question is raised by most people when they first study fuzzy set theory. There are no standard procedures to be applied in every case; anything plausible would seem to do. In particular, there are neither behavioral specifications nor canonical examples of the kind Shafer claims to be important. Zadeh would argue that a theory of imprecision should not need precise inputs, so that we should not bother too much over the exact nature of the input membership functions. If that is the case, then answers should not be very sensitive to input membership functions. In many applications, this is not the case, and indeed, sometimes answers are sensitive to just one point on a membership function.

<u>What is the meaning of the output</u>? Paralleling the uncertainty relationship between human perceptions of imprecision and the calculus of fuzzy sets is the reverse relationship: once we have computed an output fuzzy set, what do we do with it? We briefly discussed the possibility of linguistic interpretation above. This does not appear to have been a satisfactorily implemented

approach, although in part because people differ in the conclusions they draw from the same natural language statement.

In the light of these difficulties, it is not surprising that efforts should be made to assimilate fuzzy sets to some other framework of uncertainty, such as the Bayesian or Shaferian. It is difficult to do this in a natural way, however, due to the difference between imprecision and uncertainty about facts. For example, suppose Analyst A refers to an object x as "long", after having measured x exactly. There is no doubt as to x's actual length and although A may regard x as long only to a certain degree, he is not uncertain whether or not x is long. What fact then could A be uncertain of? We add three caveats: (i) if A tells a second Analyst B that x is long, then *B* may be uncertain regarding x's actual length; (ii) if A had only glanced at x, rather than measuring it, he might be uncertain (as well as imprecise) about x's actual length; (iii) we may in fact be uncertain as to whether a random English speaker would call the object "long". Nevertheless, the most natural approach is to treat this kind of uncertainty as the degree to which x (or an object of x's length) is long, rather than the chance that x is long. Put another way, these degrees are part of the meaning (denotation) of "long", and not (necessarily) a result of uncertainty about what "long" means or about the actual length of an object.

Nonetheless, it may be worthwhile exploring ways to represent imprecision in terms of other frameworks. For example, a consonant Shaferian support function obeys a calculus that closely approximates Zadeh's possibility theory. Consonant support functions seem appropriate for representing imprecision in the implications of evidence (it points to a set of nested regions where the truth could lie). And they have the advantage of a somewhat more secure normative foundation. Thus, the possibility of translating between natural language expressions and support functions might be worth exploring, despite some cost in naturalness.

Inference: <u>What are the appropriate connectives</u>? In terms of either axiomatic justification or face validity, the procedures Zadeh recommends for combining his membership functions are not unique. For example, Zadeh argues that the degree to which an element belongs to a set A_1 <u>and</u> another set A_2 should be computed by

$$\mu_{A_1 \cap A_2}(x) = \min(\mu_{A_1}(x), \mu_{A_2}(x))$$

This is clearly consistent with the requirement that if both sets are *crisp* (i.e., only takes the values 0 or 1), set membership should obey the usual rules (i.e., x A_1 A_2 if and only if x A_1 and x A_2). Note however, that this is not the only connective rule with this property. For example, the family of connectives

$$\min(\mu_{A_1}(x)\mu_{A_2}(x)^{1-\alpha}, \mu_{A_2}(x)\mu_{A_1}(x)^{1-\alpha}), \qquad 0 \leq \alpha \leq 1.$$

all have this property, where $1-\alpha$ is a power to which the membership function is raised. Zadeh chooses $\alpha = 1$; the choice of $\alpha = 0$ gives the Bayesian rule for the probability of a conjunction (namely $\mu_{A_1}(x)\mu_{A_2}(x)$). There are many other possible definitions (see Dubois and Prade, 1984).

Similarly, disjunction, negation and implication all have alternative representations, and the choice of the forms usually employed is arguable. So far as we are aware, very little research has been carried out on the implications of using different connectives on the results of a fuzzy analysis. There is, therefore, some arbitrariness in the connectives chosen by Zadeh--an arbitrariness which pervades the theory.

Plausibility of instances: The main strength of Zadeh's theory is in its ability to produce instances of reasoning that are acceptable on a case by case basis. In this regard, it has a richness and scope that no other theory even attempts to capture. In particular, it is the only theory that attempts to formalize the combination of considerations based on *similarity* (e.g., the closeness of F* to F in the above example) with more traditional considerations in inference (e.g., traditional logic or probability). In this largely uncharted domain, the (present) absence of deep normative foundations may be no disgrace.

Nonetheless, there may be cases where fuzzy logic gives implausible (or non-useful) answers. Fuzziness is concerned with what is possible, rather than

what is probable. Zadeh sees a possibility distribution as being an upper bound on a probability distribution. Articulating the possible may be important, but if many options are possible, it does not help in our search for what is probable. In practice, this point is expressed by the tendency for fuzzy sets to produce rather bland answers, giving high values of the membership function for large sets of variables. One can see some applications when this is not an obstacle to understanding, if some important options are seen to have very low or zero possibility. In general, it does present a difficulty.

Summary. Fuzzy logic is a highly flexible and versatile tool for handling imprecision. It may be applied directly to reasoning with verbal expressions or, at a higher level, to reasoning with a numerical calculus like probability theory. Unfortunately, the meaning of fuzzy measures is not always clear; and the rules for manipulating them seem to lack any deeper justification than the plausibility of the answer in a specific application.

2.3 Qualitative Theories

2.3.1 Classical logic. Only a brief mention will be given here of classical logic. Its relevance is as the traditional paradigm of analytical reasoning, dating back to the time of Aristotle and achieving maturity in twentieth century mathematical logic associated with such names as Russell, Godel, Church, and Tarski. As such, it provides a point of comparison for other theories.

Classical logic is built upon a firm axiomatic foundation of principles for reasoning from a set of premises to a conclusion. Straightforward procedures exist for checking the validity of an argument. A number of features of classical logic, however, make it clearly inadequate as the sole basis for an expert reasoning system, or as an analytical model of real-life human reasoning.

- Classical logic moves from certain premises to certain conclusions. No provision is made for reasoning in uncertain domains.
- Due to its abstract nature, there is difficulty mapping messy real-world problems into the crisp inputs required for logical analysis.
- Logical implication is very different from causal implication.

- Classical logic is not equipped to deal with causal relationships among variables.
- Classical logic is *monotonic*, i.e. the number of provable statements increases monotonically with the number of premises. In contrast, human reasoners often adopt provisional assumptions, deriving conclusions which may later be retracted when new information invalidates the assumptions.

In Section 2.2 we have seen examples of inference frameworks designed to address some of these shortcomings. Jeffreys (1939) developed his axiomatic function for Bayesian inference as an extension of classical logic to truth values intermediate between certainly-true and certainly-false. The various axiom systems for Bayesian inference are clearly modeled after those for classical logic.

Zadeh's fuzzy set theory, as we have seen, was developed to counter the second problem, which is shared by Bayesian and Shaferian theories.

The third problem was also noted in our discussion of Bayesian theory. Shafer's theory is better equipped to deal with causal links than is logic or Bayesian theory. Specifically, the link between evidence and conclusion in an argument on which a belief function is based may be a causal model according to which the evidence causes the conclusion.

The next section describes an attempt to deal with the last problem by formulating a reasoning system that reason's from "default assumptions" to conclusions which may be retracted if the assumptions on which they are based turn out to be false.

2.3.2 Non-monotonic reasoning. *Nature of the Theory*. Non-monotonic logic has its roots in the non-numeric tradition of artificial intelligence. The first application of the ideas of non-monotonic reasoning was by Stallman and Sussman (1977), and since that time the theory has generated intense interest in the artificial intelligence and expert systems communities (e.g., Doyle, 1979; McDermott and Doyle, 1980; McDermott, 1982; Reiter, 1980; Moore, 1985).

Non-monotonic logic was developed to counter the failure of traditional approaches to capture the non-monotonicity of human reasoning. Specifically,

traditional formal axiomatic logics are non-monotonic, in that the number of provable statements in the system increases monotonically in time as new axioms or premises are added to the system. In contrast, in a non-monotonic system a theorem may be retracted when new information (axioms) are introduced.

Human reasoning is commonly non-monotonic. In the face of incomplete evidence, people adopt "default assumptions," acting as if they are true until evidence arises to the contrary. For example, we might adopt a provisional assumption that there is no ECM in the area, which implies that our localization of threats is reasonably accurate. If we later discover that there is evidence of ECM, we drop the initial assumption and out confidence in the threat localizations is degraded. Human reasoners are skilled at incorporating conflicting data into existing arguments so as to achieve consistency with minimal disruption of the established system. Non-monotonic reasoning systems attempt to model this process of revising systems of belief to accommodate conflicting information.

We may contrast non-monotonic reasoning with systems based in the probability tradition, which employ numerical measures of uncertainty. In the above example, a Bayesian or Shaferian system would assign a numerical degree of support to the hypothesis that ECM is present. When further information is received, degrees of support are updated to incorporate the new information. These systems are monotonic in the sense that once a conclusion is declared certain, it cannot be retracted. Uncertainty is expressed by assigning degrees of support of less than unity to each of the uncertain hypotheses. Bayesian and Shaferian theory lack a mechanism for accepting an uncertain hypothesis once it becomes "certain enough." Conflicting evidence is regarded as stochastic (that is due to noise in the data) rather than as evidence of an incorrect model; this leads in some examples to counterintuitive results (e.g., the example of Figure 2-4). It has been argued that non-monotonic reasoning captures more fully the features of human reasoning, because of its capacity to adopt uncertain hypotheses as provisional assumptions, acting as if they were certain and deriving conclusions from them, while retaining the ability to drop them if they later turn out to be implausible. Nevertheless, non-monotonic systems suffer from the inability to distinguish degrees of certainty. The ad hoc nature of their mechanisms for belief revision (see

discussion below) is in large part attributable to this inability. There have been suggestions (Ginsberg, 1984; Cohen et. al., 1985) of means of combining degrees of belief into non-monotonic systems, but thus far none has been implemented in an expert system.

Structure of a Non-Monotonic System: Dependency Directed Backtracking. An important feature of a non-monotonic system is its mechanism for revising beliefs in the presence of new evidence. At any point in time, a non-monotonic system has a list of currently believed statements, together with a record of how these beliefs were derived. As long as new information is consistent with current beliefs, the system incorporates the new information by combining it with the currently believed statements, using its inference rules to derive new beliefs. At some time, however, new information may lead to an inference that contradicts a currently held belief. When this happens, the system must change some of its beliefs so as to achieve consistency. Because it retains a record of the proofs of each of the contradictory statements, the system need only re-examine those beliefs actually contributing to the contradiction. Thus, the system traces back through the proofs to find those beliefs upon which the contradictory inferences depend, and makes the necessary revisions to achieve consistency. This process has been labeled dependency-directed backtracking.

Dependencies in a non-monotonic system are represented by justifications of statements in terms of other statements. The primary form of justification is a data structure called a support list. A support list justification for a statement has the form

Statement # Statement (SL <inlist> <outlist>).

A statement is believed if it has a valid justification; a support list justification is valid if every statement in the inlist is believed and every statement in the outlist is not believed. We may distinguish three types of justification: premises, monotonic justifications, and non-monotonic justifications.

A premise is a statement with empty inlist and outlist. For example:

N-1 Agent testifies invasion is planned (SL () ()).

This statement is automatically regarded as IN (i.e.,.. believed) and cannot be retracted. Premises might be observational data or unquestioned general principles. As new observations accumulate, new premises may be added to the system.

A monotonic justification has a non-empty inlist and an empty outlist. For example:

N-2 Invasion is planned (SL (Agent so testifies, agent is trustworthy) ()).

This justification says that the statement is IN if all items in the inlist are IN.

Non-monotonic justifications, or assumptions, have non-empty outlists. For example:

N-3 No invasion is planned (SL () (Invasion is planned))

This statement says that N-3 is IN unless there is evidence to the contrary.

Now let us see how a non-monotonic system might handle our example. We start by adding an additional assumption

N-4 Agent is trustworthy (SL () (Agent is untrustworthy).

Let us suppose we begin with nodes N-2, N-3 and N-4 as the only items of information in the system. N-3 and N-4 are IN (our having no evidence to the contrary) and N-2 is OUT. This models the situation before we received the agent's report. Once we receive the agent's report, we add the premise N-1. This causes N-2 to move IN.

We reasoned above that N-2 and N-3 were contradictory, and responded by dropping the assumption N-3. To mimic this reasoning, the system needs to know

that these nodes are incompatible. Thus, the system needs to have another node

N-5 CONTRADICTION (SL (N-2, N-3) ()).

When N-2 moves IN, this causes the CONTRADICTION node N-5 to move in also. The presence of a CONTRADICTION mode among the currently belief statements triggers the process of dependency-directed backtracking. When a contradiction is encountered, the system searches for the set S of assumptions responsible for the contradiction. The set S will contain any assumptions in the support lists any nodes involved in deriving the argument leading to the contradiction. In this case, S will contain N-3 and N-4. Clearly, if N-3 is taken OUT, N-5 will move OUT and the contradiction will be resolved. If N-4 is OUT, then N-2 moves OUT and again the contradiction is removed.

The system now sets up a new node of the form

Statement # NOGOOD S (CP(CONTRADICTION) (S) ())

where CP is a conditional-proof justification. This CP-justification is valid if whenever S is valid, the CONTRADICTION is believed. In other words, the validity of the justification depends on the relation between the premise (S) to the conclusion (CONTRADICTION), irrespective of whether the premise is currently believed. In our example, the system would define

N-6 NOGOOD (N-3, N-4) (CP(N-5) (N-3, N-4) ()).

This CP-justification is valid because N-5 is IN whenever N-3 and N-4 are both IN. This node says that N-3 and N-4 are, taken together, "no good."

The system now has to decide which of the assumptions in S is to be dropped. A "culprit" C is selected from among those nodes in S, and the system decides to deny that assumption. Recall that to deny an assumption, the system must believe some member of the outlist of the assumption. The system does this by setting up a support list justification for some member O of the outlist of the culprit. The inlist of this justification contains all the assumptions in S except C, together with the NOGOOD node. The outlist contains all the nodes

in the outlist of C except O. Thus, the justification says that if you want to believe the other assumptions in S (other than C) and if you do not believe any other nodes in the outlist of C, then you should believe O. The result of this justification is that O is believed (provisionally), sending C OUT and resolving the contradiction. Of course, O itself may later have to be retracted as a result of another contradiction, which would lead either to belief in some other member of the outlist of C or to the retraction of some assumption other than C.

Let us return to our example, and suppose that N-3 is selected as the culprit. The outlist of N-3 has only one member, N-2. A new justification is then set up for N-2, which now appears as

N-2'	Invasion is planned	(SL(Agent so testifies, agent is trustworthy)()).
		(SL(N-6)()).

Note that if N-3 had other nodes in its outlist, the new justification would have a non-empty outlist, and would cease to be valid if one of the nodes in its outlist came IN.

It appears that N-2' can now be justified either by the agent's testimony or as an assumption required to resolve the contradiction represented by N-6. But the second justification is circular, because it was N-2 that gave rise to the contradiction in the first place. Doyle's Truth Maintenance System guards against such circularity by designating some justifications as "well-founded" and others as not.

Feasibility. Dependency directed backtracking is a species of discrete relaxation (like Walz filtering, as described in Cohen and Feigenbaum, 1982). It seeks a consistent allocation of truth values across a set of statements, by utilizing local consistency constraints between pairs of statements, rather than by exhaustive search through the space of all possibilities. Thus, a high level of computational efficiency can be achieved.

To make this efficiency possible, however, in non-monotonic systems, the traces of proofs are retained, even though the premises utilized by the proof,

and the statement that was proved, may (temporarily) be judged invalid or OUT. Therefore, if the premises become valid or IN at some later time, the work of rediscovering the proof need not be repeated. The justifications consume space in memory, and the tradeoff is therefore made between memory storage and the processing overhead of regenerating proofs on the fly.

Face validity. Implementations of non-monotonic reasoning revise beliefs so as to arrive at a consistent overall system of beliefs in the face of a contradiction. But they provide only a very limited capability for deciding among alternative possible revisions. The selection of an assumption as the "culprit," and the selection of a member of its outlist to be assumed as true, are both highly arbitrary. Some control information is implicit in the ordering of nodes in the outlist of statement 5; i.e., if 5 is to be rejected, the system will then assume the truth of members of numbers in the outlist in the order shown. But (a) this is insufficient to remove all ambiguities, and (b) it makes control information implicit rather than explicit, hence, difficult to evaluate or modify.

Plausibility of instances: Conflicting evidence. An often voiced criticism of non-monotonic reasoning is that uncertainty calculi (e.g., Bayesian, Shaferian, or fuzzy) can do the same job better.

Although we are convinced of the value of numerical representations of uncertainty, we will argue that there is an important role of non-monotonic reasoning (1) in drawing implications for the validity of one argument or line of reasoning from another, even where they are independent, and (2) in reasoning about the application of the uncertainty calculus itself.

The basic idea of (1) is the following: Suppose we have two independent lines of reasoning, A and B, with regard to the same sets of hypotheses. Each line of reasoning depends on certain data and certain assumptions, as illustrated in Figure 2-6. In Argument A, the impact of Data 1 and Data 2 depends on the acceptance of Assumption 1; for Argument B, the impact of Data 3 and Data 4 depends on Assumption 2.

What happens when A and B support conflicting hypotheses? In a non-monotonic system, the set of assumptions that contributed to the contradiction are iden-

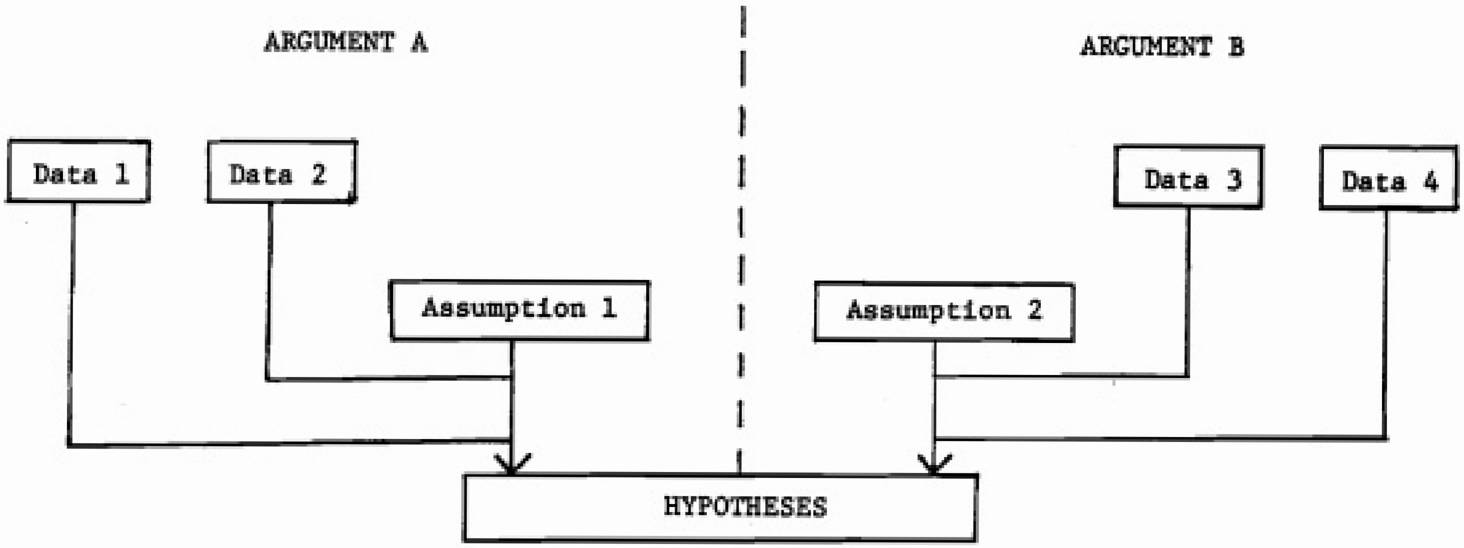

ARGUMENT A
ARGUMENT B
Data 1
Data 2
Data 3
Data 4
Assumption 1
Assumption 2
HYPOTHESES

Figure 2-6

tified and declared inconsistent (as a set). Then a selected member of this set is rejected, by producing a justification (itself an assumption) for a member of its <u>out</u>list. As a result, at least one of the two arguments fails (or has a different conclusion), and consistency is restored.

The key point here is that conflict between A and B causes the system to *reach inside* each of the arguments. Conflict resolution is a process of reasoning about knowledge: what are the weakest links in each line of reasoning? where would revision accomplish the most?

Consider, on the other hand, how an uncertainty calculus such as Shafer's would handle this problem. We examined the issue of conflict resolution, in the context of belief function theory, in some detail in Section 2.2.2. There we found that, depending on the degree of conflict, and on the existence and degree of discounting for the two arguments, we could have: (a) an indeterminate result (if there is no non-empty intersection between possible meanings of the two arguments), (b) exclusive support for hypotheses in the intersection of meanings (if there is no discounting), or (c) strong support for each of the two conflicting conclusions). None of these alternatives examines the sources of the conflict and seeks insights regarding its causes. Adjustments of discount rates in the light of conflict are likely, moreover, to be invalid in the absence of some exploration of reasons for the adjustment.

Nonetheless, non-monotonic systems as presently constituted are inadequate in a number of ways. Problems are chiefly attributable to their *exactness*, on two levels. For example, non-monotonic systems provide a way of reasoning with incomplete information, i.e., by adopting assumptions, tracing their consequences, and revising them if they lead to an inconsistency. But they provide no *measure* of the degree of incompleteness in the support for a belief, and no concept of degree of conflict. As we have already noted, a measure of this sort seems essential in selecting among alternative possible revisions.

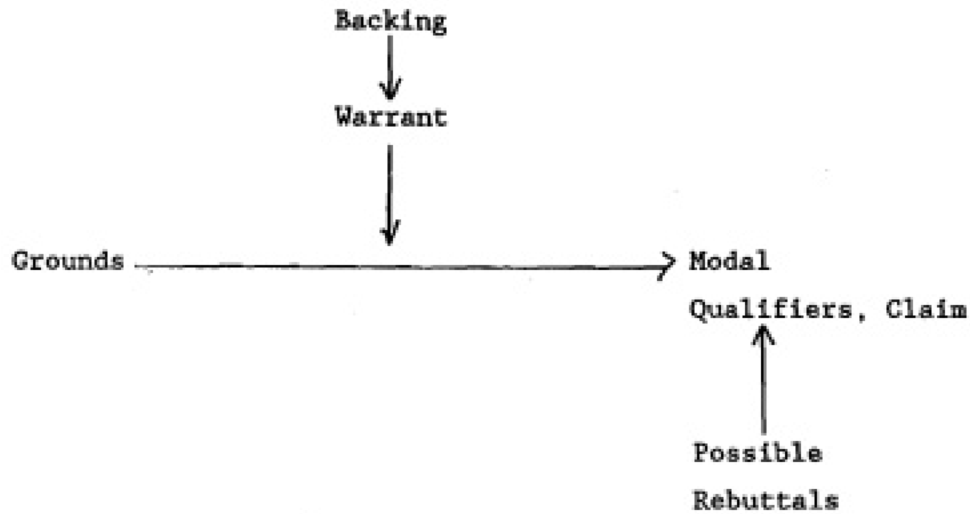
On a second level, the statements whose truth or falsity is adjudicated are themselves exact. However, there is no reason why similar principles of qualitative reasoning might not be applied to probabilistic or imprecise constraints and data. The need for such a "meta-reasoning" capability is the

chief conclusion of our comments in earlier discussions of Bayesian and Shaferian calculi. In our view, non-monotonic logic may have its most convincing application at a higher level, in controlling the application of an uncertainty calculus itself. Assumptions of more than one sort--about the quality of uncertainty assessments, about the independence of evidential arguments, and about the validity of steps in an argument--are inescapable in the application of such a calculus. Most of these assumptions are not easily represented in the language of the calculus itself. Hence, non-monotonic reasoning may be the appropriate tool for keeping track of assumptions and revising them when they lead to anomalous results. As such, it may be the key to a truly "intelligent" or flexible application of those models. It is to this possibility that we turn in Section 3.

Summary. Non-monotonic logic is a computationally efficient method for reasoning with incomplete information, i.e., for adopting assumptions and revising them in the face of conflicting data. Statements are associated not with numerical indices of uncertainty, as in the other theories we have examined, but with reasons. Certain statements (called assumptions) may be accepted in the absence of positive support, as long as certain other beliefs have not been disproven. Non-monotonic logic provides a natural method for revising beliefs within independent lines of reasoning when they lead to conflicting conclusions. Unfortunately, validity is diminished by the arbitrariness of its procedures for selecting among alternative possible belief revisions. We argue that the most useful application of non-monotonic reasoning may be as a control process for the application of an uncertainty calculus.

2.3.3 <u>Toulmin's model of logic</u>. The motivation of Toulmin's *Uses of Argument* (1958) is to turn away from the highly abstract character of traditional logic; to examine actual methods of reasoning in different substantive areas, such as law and medicine; and to develop a theory of logic capable of capturing the rich variety of methods that exist. In the preface he states, "the intentions of this book are radical." He rejects as confused the "conception of 'deductive' inference which many recent philosophers [and, we may add, AI researchers] have accepted without hesitation as impeccable."

The basic framework of an argument, according to Toulmin, is as follows (Toulmin, et al., 1978):



A claim, or conclusion whose merits we are seeking to establish, is supported by grounds, or evidence. The basis of this support is the existence of a warrant that states the general connection between grounds and conclusion: e.g., a rule of the form, if this type of ground, then this type of conclusion. The backing provides an explanation of why the warrant is regarded as reliable, i.e., it provides evidence (theoretical or empirical) for the existence of a connection between ground and claim. Modal qualifiers weaken or strengthen the validity of the claim. Possible rebuttals deactivate the link between grounds and claim by asserting conditions under which the warrant is invalid. A way of reading this structure is: Grounds, so Qualified Claim, unless Rebuttal, since Warrant, on account of Backing.

Toulmin finds serious fault with purely analytical or logical arguments. In such arguments (as contrasted with a substantial argument) the backing includes the information conveyed in the conclusion. As a result, of course, the backing can be no more certain than the conclusion itself. In ordinary arguments, by contrast, "we seek to establish conclusions about which we are not entirely confident by relating them back to other information about which we have greater assurance." Moreover, the certainty of the conclusion (e.g., a prediction of a future event) is seldom *logically* entailed by the grounds and backing (e.g., past observations of situations like the present one); it is merely made more plausible (and of course, rebuttals may always turn up to

reduce its plausibility). Toulmin concludes, "it begins to be a little doubtful whether any genuine, practical argument could *ever* be properly analytic."

In particular, Toulmin points to weaknesses in the use of the logical term 'universal premise.' His illustration (p. 115) highlights the weakness.

Jack is club-footed.
All club-footed men have difficulty in walking.
So, Jack has difficulty in walking.

In a logical pattern of analysis, the general statement 'All...' is construed as an abstract inference-warrant for deriving the conclusion from the evidence. In a real argument, we would never supply such a statement as backing for a conclusion. Our actual backing might be that all club-footed men *observed* by us have had difficulty walking (an empirical basis), or that the nature of club foot suggests difficulty in walking (a more theoretical backing). Toulmin concludes (p. 117), "the form 'All A's are B's' occurs in practical argument much less than one would suppose from logic textbooks."

According to Toulmin (p. 143), "the traditional pattern of analysis has two serious defects. It is always liable to lead us to pay too little attention to the differences between the different modes of criticism to which arguments are subject". In addition, the traditional pattern has the effect of "obscuring the differences between different fields of arguments, and the sorts of warrant and backing appropriate to these fields."

On probability, Toulmin rejects the subjectivist's probability as the degree of belief on the basis that this is incompatible with the requirement that estimates of probability be reliable. He also rejects the objectivist's definition of probability in terms of frequencies, on the basis that such a definition confuses the meaning of probability (i.e., as a qualification of a conclusion) with the reasons for regarding the event as probable (i.e., the observed frequencies). In fact, he contends that, "the attempt to find some 'thing', in terms of which we can analyze the solitary word 'probability' and which all probability-statements whatever can be thought of as really being about, turns out to be a mistake" (p. 70). He defines probability as a modal qualifier asserting, "whether backed by mathematical calculations or no, the

characteristic function of our particular, practical probability-statements is to present *guarded* or *qualified* assertions and conclusions" (p. 93).

Toulmin's framework bears some important resemblances to non-monotonic logic. Both depart from traditional logic by providing for a process in which conclusions are accepted *unless* other propositions (members of the <u>out</u>list; rebuttals) turn out to be true. There are two important differences: (1) Toulmin proposes a highly differentiated knowledge structure, in which grounds, warrant, backing, conclusion, and rebuttals are distinguished, while non-monotonic logic proposes an essentially homogeneous, undifferentiated knowledge structure; (2) Toulmin provides for graded or qualified acceptance of conclusions.

In Section 3, we shall use Toulmin's basic framework as a starting point for a model of argumentation from evidence to conclusion on which a Shaferian belief function is based. We shall see how, when conflict occurs, a process of non-monotonic reasoning can "reach inside" the arguments, exploring potential rebuttals, and leading to revision (i.e., discounting) of the component belief functions and reduction of the conflict.

2.3.4 <u>Theory of endorsements</u>. Paul Cohen's (1985) theory of endorsements is another descendant of the AI-based logic tradition. Although non-numeric in character, there is an interesting commonality in motivation with Shafer's theory. Both methods focus on the validity of arguments that purport to establish a conclusion based on evidence. For Shafer, however, one's belief about such an argument can be adequately summarized in a numerical measure, the belief function, i.e., the likelihood that the evidence proves the hypothesis. To Cohen, by contrast, it seems unnatural to assess the strength of an argument without actually examining the argument in detail. They theory of endorsements provides a consistent format for representing such arguments.

In Paul Cohen's theory of endorsements, evidence is represented not by numerical measures of degree of belief, but by symbolic endorsements. A given proposition is associated with a "ledger" of confirming and disconfirming evidence. Each item of evidence, in turn, is associated with a set of positive and negative "endorsements," which state grounds for believing or disbelieving a link between that evidence and the hypothesis. Finally, the

theory contains rules for ranking different types of endorsements, for determining when they qualify a hypothesis for acceptance, and for resolving conflicts.

Cohen's theory has been implemented in a prototype system called SOLOMON (Cohen, 1985). The user of Cohen's system supplies primary data and inference rules with endorsements (e.g., a rule may be endorsed MAYBE-TOO-GENERAL). Endorsements of a rule and the propositions to which it is applied propagate to the conclusion of the rule (and, as noted above, can be thought of as endorsements for the *linkage* between the evidence and the conclusion). The system must be supplied with criteria for when a proposition is adequately (for a particular purpose) endorsed; these criteria depend on the goal as well as on the endorsements for the proposition.

It is worth noting that Cohen's concept of an endorsement encompasses a variety of distinguishable elements of Toulmin's framework: i.e., warrant, backing, and rebuttals may all serve as (positive or negative) endorsements affecting the link between ground (evidence) and conclusion.

Cohen's approach has a unique simplicity and transparency, and may capture a significant aspect of actual reasoning (the dependence of belief on qualitative facts about the available evidence). Nevertheless, as with the other theories reviewed here, the utility of Cohen's theory depends on several unresolved issues. First, the ranking of endorsements is entirely qualitative. Cohen expresses concern that for some proposes it might be desirable to specify numerical measures of the strength of endorsements, but seems to regard this as incompatible with the symbolic reasoning tradition of AI. Numerical measures of strength of endorsements, coupled with a mechanism for combining them, would provide a resolution of the second problem: the *ad hoc* nature of the mechanism for ranking endorsements. Cohen assumes that the system is supplied with rankings for individual endorsements, but there exists only an *ad hoc* mechanism for ranking groups of endorsements. Thus, the decision of whether one proposition is better endorsed than another is to some degree arbitrary, and the rules can be insufficiently powerful to derive a conclusion. Third, endorsements are tokens (to the system). The rich associations a human would bring to an endorsement of, e.g., MAYBE-TOO-GENERAL, are opaque to Cohen's system. (It is interesting that Cohen's system was

developed in response to a perception of the opacity of numerical probability judgments, but his system suffers to some extent from the same problem.) Finally, Cohen's theory, like the numerically based theories reviewed in Section 2.2, would benefit from a "meta-reasoning" capacity for re-evaluating endorsements as an argument proceeds.

2.4 Probability/Logic Syntheses

2.4.1 Model based on Toulmin's framework. Lagomasino and Sage (1985) present a framework for imprecise inference that purports to combine Toulmin's logic of reasoning and the calculus of probability. In fact, we would argue that their use of Toulmin is quite incidental to their basic approach. A better characterization is that Lagomasino and Sage attempt to probabilify traditional logical relationships.

Lagomasino and Sage claim to use Toulmin's model of argumentation to frame the relations among events, and to structure an inference model. In particular, the relationship between two events, grounds D and claim C, are represented as:

$$\begin{array}{c} W = (D \rightarrow C) \\ \downarrow \\ D \rightarrow X \rightarrow C \\ \uparrow \\ R = (\bar{D} \rightarrow C,\ \bar{D},\ \bar{C},\ \ldots). \end{array}$$

[Their use of the term rebuttal to include the negation of grounds or claim appears at odds with Toulmin's (1958) definition of rebuttal as "indicating circumstances in which the general authority of the warrant would have to be set aside" (p. 101).]

Probabilities serve as modal qualifiers and the calculus of probability is used to combine or aggregate assessments. Within this structure, both uncertainty and imprecision about uncertainty are represented. Uncertainty about the validity of a proposition or strength of a claim is presented as a probability. Imprecision about uncertainty is represented as ranges on

probabilities. [Toulmin (1958) uses probability only as a modal qualifier on claims.]

Lagomasino and Sage derive a set of consistent relationship equations (CRE) based on logically consistent relationships among claims, grounds, and warrants (collectively called premises) and possible rebuttals and the rules of probability. [Again, the approach diverges in spirit from Toulmin (1958), who dismisses as trivial the notion of formal validity by noting, "provided that the correct warrant is employed, any argument can be expressed in the form 'Data [Grounds in his later terminology]; Warrant; so Conclusion' and so becomes formally valid" (p. 119).] The following set of linear, independent equations and inequalities is the set of CREs for the above:

$$
\begin{aligned}
P(D \rightarrow C) + P(D \rightarrow \bar{C}) + P(D) &= 2 \\
P(D \rightarrow \bar{C}) + P(\bar{D} \rightarrow \bar{C}) + P(C) &= 2 \\
P(\bar{D} \rightarrow C) + P(D \rightarrow C) - P(C) &= 1 \\
0 \leq P(\cdot) \leq 1.&
\end{aligned}
$$

This framework is used to derive probability statements concerning any premise or rebuttal by solving two linear programs. The CREs are the set of constraints, and the objective functions are determined by the premise or rebuttal of interest, namely min $P(\cdot)$ and max $P(\cdot)$.

As they stand, the basic sets of CREs do not say anything interesting. That is, each $P(\cdot)$ has a range of 0 to 1. These ranges are narrowed only by the addition of information.

Information is represented in this system as additional constraints. The following are examples of constraints that might be provided by information:

$$P(\bar{D} \rightarrow \bar{C}) < 2P(D \rightarrow \bar{C}),$$

$$P(D \rightarrow C) > P(\bar{C} \rightarrow D),$$

$$P(D) = .75.$$

In some cases information might contain a term that does not appear in the canonical representation of CREs. The second example above, which contains the term $P(\bar{C} \rightarrow D)$, is such a case. Such relationships can be converted to

canonical form using equivalence relationships. In the example, the equivalence $\bar{C} \rightarrow D \equiv \bar{D} \rightarrow C$ is invoked so that $P(\bar{C} \rightarrow D) = P(\bar{D} \rightarrow C)$. A similar procedure can be used to transform information that is provided in other ways. An important case is conditional probability statements. For example, the statement $P(D|C) > .6$ can be transformed to canonical form as follows:

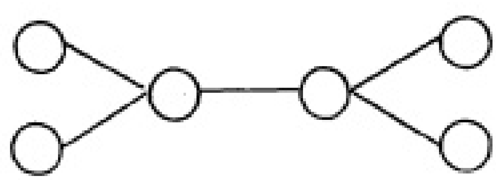
$$P(D|C) > .6$$

$$\frac{P(D \cap C)}{P(C)} > .6$$

$$\frac{1 - P(D \rightarrow \bar{C})}{P(C)} > .6$$

$$P(D \rightarrow \bar{C}) + .6P(C) < 1.$$

The model can be expanded to represent a whole network of events. Pictorially, Lagomasino and Sage show such a network as,



but from the generality of their framework and discussion, the method does not appear restricted to a spanning tree. The link between any pair of nodes in the network may consist of a subset of basic premises and possible rebuttals. As stated, links appear to be quite general (presumably some node's claim could become another node's rebuttal), and thus the approach appears to lack the intrinsic structures of Toulmin's logic. Each relationship is modeled as a set of CREs, and the set of all CREs and additional information constraints are constraints in the linear program. This system of relationships may be solved to determine the range of probability of any factor of interest. If the linear program has no feasible solution, then those relationships specified are logically inconsistent. (This suggests an extension of the method using goal programming techniques, but a logical basis for such an extension is not apparent.)

Lagomasino and Sage do not explain how the informational relationships should be assessed or estimated, or exactly how new information changes a set of

constraints. Moreover, information that loosens as well as tightens bounds (i.e., non-monotonic reasoning about probabilities) is presumably possible, but Lagomasino and Sage are silent on how this could occur in their framework.

Nor is the process of specifying nodes in the network ever defined precisely. Lagomasino and Sage state that structuring a model involves "the specification of alternative hypotheses at each node" and that "the set of hypotheses under consideration at each node should be mutually exclusive and exhaustive." However, hypotheses are also limited to propositions that obey the consistency relational equations. It is unclear how this constraint on assessment can be enforced.

According to Lagomasino and Sage, the method allows for information to be encoded about both causal (e.g., P(D→C)) and diagnostic (e.g., P(D|C)) reasoning. This claim highly dubious and represents, we think, the most serious weakness in this approach. P(D→C) *cannot* plausibly by construed as the probability (or strength) of a causal link between D and C as long as "D→C" is interpreted within traditional logic (as the authors clearly intend). Within traditional logic "D→C" is true unless D and C are both true (this is the interpretation used by Lagomasino and Sage in the derivation described above regarding P(D|C) and P(D→C)). But this is far weaker than a causal connection: "If the moon is made of green cheese, then the threat is an SA-4" would be true in traditional logic, since the antecedent is false; yet clearly there is no causal connection. A warrant construed in this way is quite trivial: when the antecedent is true, the warrant merely states that the conclusion is true. It gives no indication of any physically real connection between the two.

An alternative interpretation of "D→C" is as an implicit universal generalization, i.e., all instances of D are also instances of C. This runs into the objections broached by Toulmin. In particular, a single counterexample (i.e., a case of D and not-C) is sufficient to establish the falsity of such a generalization; i.e., P(D→C) would be zero. Yet we often assert the existence of causal relations (e.g., "the baseball's hitting the window caused it to break") even when the relationship is subject to exceptions (some baseballs would not have broken some windows).

In the light of these problems, two broad courses of action are available: (1) we can interpret "$\rightarrow$" outside of classical logic, e.g., in terms of modal logics for causality. In this approach, $D \rightarrow C$ is true (i.e., D causes C), for example, only if C is true in all the physically possible worlds where D is true. Perhaps the *degree* to which D causes C is the percentage of D worlds where C is also true. This option involves enormous difficulties computationally (i.e., in specifying CRE's within a modal framework) and semantically (i.e., in defining the notion of a "possible" world precisely enough so they can be counted). (2) A simpler option is to take a causal (or other theoretical) link as a basic unanalyzed notion, and to assess the probability of its existence. This is essentially Shafer's approach in his notion of evidential support. Thus, $m_E(H)$ can be interpreted (with qualifications discussed in Section 2.2.2 above) as the chance that the evidence E proves or establishes the hypothesis H. In cases of causal reasoning, this is the chance that E causes H or that H causes E. (For example, the reliability of a witness who claims that artillery is present is simply the probability that his testimony was in an appropriate causal relation to the presence of artillery.)

In the framework to be described below, we in essence adopt course (2). However, we supplement Shafer's simple representation by an explicit analysis of the basis of the alleged evidential link in each argument: i.e., its backing and its possible rebuttals.

2.4.2 <u>Nilsson's probabilistic logic</u>. Nilsson (1984) presents an approach that, on the surface, appears very similar to that of Lagomasino and Sage. Nilsson proposes a method for characterizing the truth-values of first-order sentences as probabilities. The method is applicable to "any logical system for which the consistency of a finite set of sentences can be established." This method is presented as a generalization of classical first-order logic that is "appropriate for representing and reasoning with uncertain knowledge."

Nilsson starts by specifying a logical sentence whose truth values are of interest. These could be any conjunction of sentences of first-order logic. For example, a sentence could be:

$$S = \{(\exists y)A(y),\ (\forall x)[A(x) \supset B(x)\ ],\ (\forall z)B(z)\}.$$

The truth-value of any one of the three components of this sentence is bounded by logical consistency relationships. For example, all three components could be true; this is logically consistent. However, the three components could not all be false; this is inconsistent. Note that this bounding is on the combination of truth-value for all components of the sentence, not for any individual component. Indeed, in the example any component could be true or false (value of 0 or 1); it is only combinations that are prohibited.

Each permissible combination of truth-values represents a "possible world," that is, a possible combination of true and false components. If the truth or falsity of each component is represented by the number 1 or 0 respectively, then a possible world can be represented as a three-dimensional vector of zeros and ones for a permissible state. In the example above, the following five vectors represent all possible worlds:

$$[1,1,1]$$
$$[1,0,1]$$
$$[1,0,0]$$
$$[0,1,1]$$
$$[0,1,0].$$

If each component of the sentence is thought of as a dimension in three-space, then possible worlds are represented as five points in that space.

Nilsson next generalizes the interpretation of the vector by allowing probabilistic "smearing" over worlds. This is done by allowing probability distributions over different worlds and by constraining these probabilities to be logically "permissible." The implication of the definition of probabilistically permissible is to constrain probabilities to be within the convex region bounded by the set of possible worlds as defined above. This leads to the following, rather tortured, interpretation of a probability of a component of a sentence:

> the probability of a component is the sum of probabilities of all possible worlds in which it is true.

Since consistency is a criterion that rarely determines probability uniquely,

Nilsson investigates additional techniques. He both solves for "maximum entropy" probabilities and those produced by geometric projection. Neither method is provided with a basis or defended. This step might take place after the permissible region is reduced by additional constraints on the probability values of sentences. No mention is made of the source of these additional constraints. So, the output of Nilsson's model is either a description of a region of permissible probabilities or a probability that is determined by an *ad hoc* method, although one could presumably assess probabilities of "possible worlds" to derive the desired probabilities (ignoring the assessment problem).

The principal difficulty of Nilsson's approach, from the present viewpoint, is that (like Lagomasino and Sage) it fails to capture true causal, or other evidential, relationships. As noted above, these are not well represented in the first-order predicate calculus, and it is not clear how effectively Nilsson's method could be extended to handle consistency constraints among sentences in a modal logic. In any case, it is clear that the assessment task would be enormously complicated (e.g., by the introduction of possible worlds containing sets of possible worlds).

3.0 AN ADAPTIVE PROBABILISTIC INFERENCE FRAMEWORK

This section describes an innovative inference framework, for use in expert systems. The framework was developed to address some of the shortcomings of current approaches to reasoning in uncertain domains.

Human experts typically use an iterative process of reassessment and revision when they reason in complex domains characterized by uncertainty. One or more models are tentatively adopted (usually requiring assumptions that are, at best, only approximately satisfied) and conclusions are derived. The plausibility of the results is assessed, by testing model results against intuition or against the results of other models. Sometimes, in addition, the model makes predictions which can be tested against actual observations. When results of an analysis meet such tests of plausibility, confidence in model assumptions is enhanced; otherwise, the human analyst searches for ways to relax or change model assumptions to achieve more acceptable results.

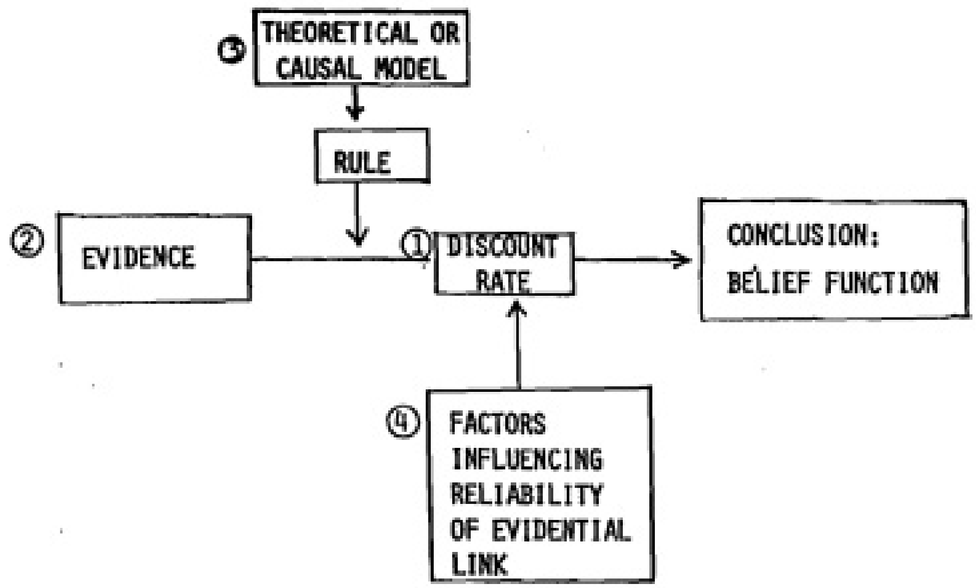
Current approaches to expert systems' reasoning under uncertainty, however, fail to capture this iterative revision process. Usually, some form of probabilistic model (e.g., Bayes, Shafer, or certainty theory) is encapsulated within the modular rules used by the system in reasoning. No provision is made for altering the probabilistic model to account for the extent to which results confirm or disconfirm model expectations. In many of these systems, moreover, there is no explicit representation of the completeness or reliability of a probabilistic argument--of the extent to which the analysis is "shiftable" with new evidence.

Another problematic feature of current expert systems is the confounding of knowledge about uncertainty with utility, or knowledge about preferences. For example, the MYCIN system handles a disease it considers serious (a utility consideration) by increasing its certainty factor (acting "as if" it is more probable than warranted by the evidence). Another common tactic is to embed utility considerations in the ordering of rule application. Such confounding makes it very difficult to maintain a knowledge base in the face of independently shifting preferences and beliefs and to communicate system reasoning to users.

Our inference framework is designed to capture important features of the iterative revision process characteristic of human reasoning. The framework takes explicit account of the "shiftability" of model assumptions, searching for potential revisions among those arguments identified as least reliable. Uncertainty is kept separate from preferences, allowing for greater normative justification of system results, for more informative user/system interaction, and for rapid adaptation to changes in system goals (this last feature being especially important in time-stressed military environments).

Figure 3-1 illustrates the representation of a single evidential argument within our reasoning framework. The representation is based on Toulmin's (Section 2.3.3) proposed model of an argument. The evidence corresponds to Toulmin's *grounds*. The *claim* (the conclusion in Figure 3-1) is linked to the grounds through the *warrant* (the rule), with *backing* provided by a causal or other theoretical model. In our framework, however, the conclusion is not a definite hypothesis, but rather a belief function which represents the system's state of uncertainty about the range of possible hypotheses. Thus, the rule links evidence to a belief function over possible hypotheses. This fits with Toulmin's conception of the role of probability as *modal qualifier* of a claim--the belief function represents a qualified (by a belief function) claim, linked to the evidence through a rule for computing the belief function, with the rule in turn backed by a causal or other theoretical model. Finally, Toulmin's framework allows for representing the reliability of the evidence, through what he calls *possible rebuttals*. In Figure 3-1, the possible rebuttals act to discount the belief function. In Section 2.2.2, in our discussion of belief functions, we saw that discounting of belief functions was a means of incorporating the judgment that there was some chance that the evidence and the hypothesis were not linked, i.e., that the evidential link was invalid due to some deactivating factor.

As shown in Figure 3-1, this inference framework has the advantages of Shafer's belief function theory: in providing a measure of the reliability of evidential arguments, in permitting modular analyses of separate lines of argument, and in the possible use of Bayesian (as well as other) types of models as special cases.



BELIEF FUNCTION REPRESENTATION

① PROVIDES MEASURE OF RELIABILITY OF EVIDENTIAL ARGUMENT

② PERMITS MODULAR ANALYSES BASED ON SUBSETS OF EVIDENCE

③ CAN USE "BAYESIAN" MODELS WHERE APPROPRIATE AS A SPECIAL CASE

IN ADDITION:

④ FACTORS INFLUENCING RELIABILITY OF ARGUMENT ARE MADE EXPLICIT

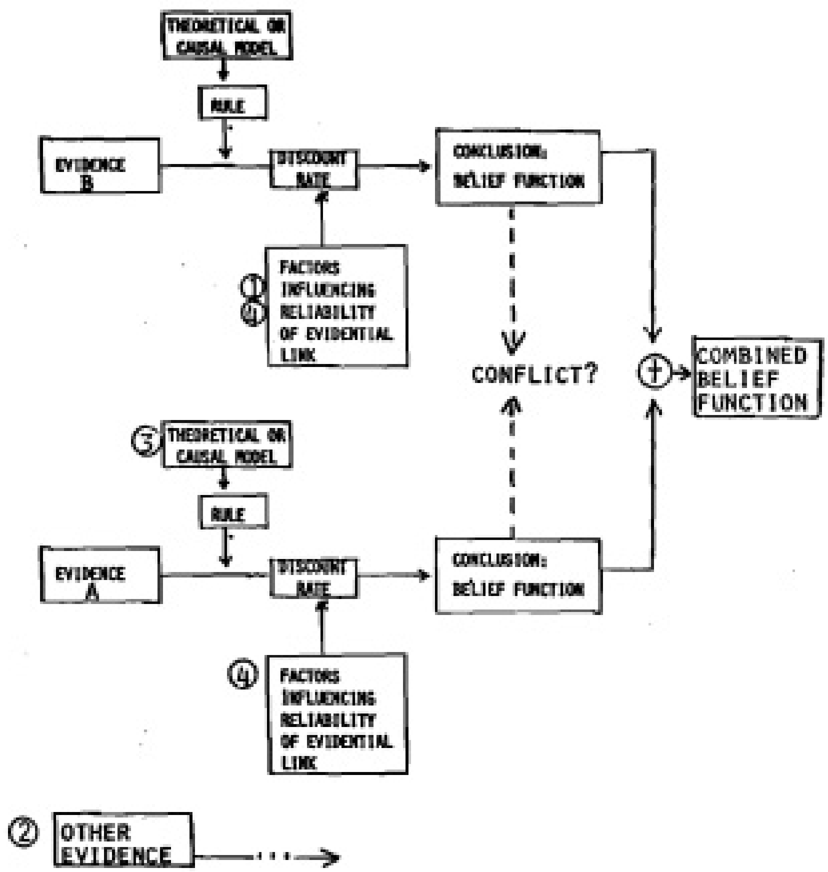
Figure 3-1

A crucial additional feature of our system, however, is that belief functions are not represented as "black boxes;" the system is provided with a frame (Figure 3-1) representing the basis for computing the belief function, together with knowledge of the factors which could discredit the link between the evidence and conclusion. Thus, the system has access not only to numerical measures of uncertainty, but to the structure of the arguments on which these measures are based. This feature provides the potential for "reaching inside" an argument and altering the resultant belief function, the alteration being based on the firmness of the components of the argument.

Figure 3-2 illustrates the combination of two arguments. The belief function representation provides for a straightforward means of combination: the application of Dempster's Rule. When the two arguments are in basic agreement, confirming each other, the inference procedure ends with the application of Dempster's Rule. However, it is possible that the arguments are in conflict, that they assign significant belief to mutually exclusive conclusions. Such a situation was discussed in Section 2.2.2, and illustrated in Figure 2-3. When such conflict occurs, the system (as would a human expert) takes it as evidence that one or more of the component arguments may be flawed; and it sets out to determine where the flaw is. As shown in Figure 3-2, the process of conflict resolution in this system can involve the application of different strategies across several different stages.

(1) The first step of conflict resolution is to search for information that may discredit one of the component arguments. Thus, the system tries to obtain information about the factors influencing the discount rate; this search is prioritized by balancing the cost of information search against the potential benefit of the information in conflict reduction. The result of the information search may be to increase belief in the presence of a factor discrediting one of the arguments; if this results in lowering the conflict to an acceptable level, the conflict resolution process is concluded.

(2) As a second step, however, the system may seek additional independent evidence related to the conclusion. This typically will result in an *increase* in conflict, but it may provide insight into which component argument is flawed (by supporting only one of the original conflicting arguments). Then,



METHODS FOR HANDLING CONFLICT:

1. SEEK INFORMATION THAT WOULD DISCREDIT ONE OF THE CONFLICTING ARGUMENTS

[2. SEEK ADDITIONAL INDEPENDENT EVIDENCE]

[3. REVISE THEORETICAL BASIS OF ONE OF THE CONFLICTING ARGUMENTS]

4. DISCOUNT ALL CONTRIBUTING ARGUMENTS

Figure 3-2: Adaptive Revision of Beliefs

by looping back to step (1), the system utilizes this additional information to reprioritize its search for the presence of discount factors.

(3) Thirdly, the system may explore modifications in the theoretical basis for one of the conflicting arguments. For example, the system may decide to modify the causal model underlying one of the inference rules, or it may decide that the arguments are not based on independent evidence, invalidating the applicability of Dempster's Rule. Clearly, such modifications are at a higher level than those discussed before, and require a system with a high-level adaptive capacity for altering the structure of its own reasoning processes in the face of unanticipated observations.

(4) Finally, if significant conflict remains, the system discounts *all* component arguments by an amount reflecting their contribution to the conflict. This reflects the conclusion that *some* element in at least one of these arguments is flawed, but that insufficient data are available (at an acceptable cost) to identify the flaw precisely.

In the demonstration system (Section 4 below), steps (1) and (4) above have been implemented.

The process of discounting belief functions when conflicting evidence is encountered is non-monotonic in character, and possesses important parallels to Doyle's (1979) non-monotonic logic, discussed above in Section 2.3.2. In a strict Shaferian system, the *input* belief functions remain fixed throughout the analysis, and combination of these functions by Dempster's Rule after the addition of new evidence always *reduces* the amount of mass allocated to the universal set. Yet Shafer himself responded to the example of Figure 2-4, where the existence of conflict resulted in a counter-intuitive result, by proposing a non-monotonic revision (discounting) of the input belief functions. Our framework provides a mechanism for implementing this non-monotonic process within an expert system. Belief functions are represented as based on assumptions (for example, until evidence to the contrary is obtained, the system acts as if a particular discount factor is absent). When conflict among belief functions is observed (conflict being analogous to Doyle's system encountering a contradiction among sentences), the system searches for a "culprit assumption" (e.g., the absence of the above-mentioned

discount factor) and looks for evidence to discredit the assumption through a test which might establish the presence of the factor. The result is a modification of the assumptions leading to one of the belief functions, and hence a discounting of that belief function and a reduction of conflict. It is worthy of note that in our system (unlike Doyle's) the prioritization of the search for "culprit" assumptions is made explicit, and is based on a benefit-cost tradeoff. Moreover, if a revision in assumptions cannot be justified (by the outcome of some test), the revision does not take place, and the system uses the device of across-the-board discounting (step (4)) to represent the overall loss in confidence in its system of beliefs.

4.0 APPLICATION TO A PROTOTYPE ADAPTIVE ROUTE REPLANNING (ARR) SYSTEM

4.1 Implementation

Our inference framework has been implemented in a small-scale prototype system designed to support pilots on deep interdiction or offensive counterair missions. The focus of the demonstration software is in-flight route replanning in the face of strategic pop-up threats, i.e., threats which are discovered at sufficient ranges to permit time for rerouting the aircraft (in contrast, for example, to the immediate evasive action required against an airborne missile). Further, the main focus among strategic threats is surface-to-air missile sites or artillery.

The Adaptive Route Replanner (ARR) is assumed to begin its mission with prior information (represented by a belief function) about the location of a particular surface-based anti-air threat. During flight, the system is notified of a second threat localization (from a SAR signal), which may be more or less distant from the likely location of the first threat. This second piece of information is likewise represented by a belief function for the location of the threat. As part of its inference task, the system must assign degrees of belief among three possibilities: (1) the two belief functions represent the same threat, in the same location; (2) the original threat has moved to a new location; and (3) the second signal comes from an entirely new threat, previously undetected.

To perform this task, the ARR utilizes its knowledge about the original threat location and the location of the new signal, as well as general information such as how far threats can move, how thorough the prior area intelligence was (and therefore, how likely to have missed a threat), and how far from the original threat a second threat is likely to be. (Other kinds of information, e.g., characteristics of the SAR signal that might help to identify the type of threat and establish whether it is the same or different as the original threat, would be included in a full-scale operational system. Due to the constraints of a limited Phase I effort, it was decided to incorporate only a

small subset of potentially available information sufficient to illustrate the inference mechanisms.)

Each of the above pieces of evidence is represented as a belief function, as described in Section 2.2.2. The system operates on these belief functions in three passes. (I) *Forward-chaining combination of belief functions using Dempster's Rule*. The result is a belief function over the three possibilities--unchanged, moved, or different--as well as a belief function over the location of the threat(s) under each of the possibilities. The analysis ends here if there is no significant conflict in the resulting belief function; otherwise, a second pass is taken. (II) *Prioritization and (possible) performance of tests*. In the second pass, the system decides on an action to take (e.g., test for ECM in the area) that might discredit one of the belief functions and result in a lessening of conflict. If the test is performed, and if a non-monotonic process of discounting occurs based on the test result, then the combination of belief functions is recomputed. Again, analysis ends if results are satisfactory. (III) *Across-the-board discounting of all arguments*. Otherwise, all component arguments are discounted based on their contribution to the conflict, and Dempster's Rule is again recomputed.

After arriving at a satisfactory inference with respect to threat classification, ARR derives the action implications of the inference. Specifically, it combines its beliefs with regard to whether the threat is unchanged, moved, or new with its knowledge of the danger contours associated with the threats, and, based on this information, evaluates several candidate routes and selects the best. (The present implementation does not generate routes, nor does it compute danger contours from more basic information. These functions have been taken as "black box" pieces of its knowledge base, in the initial phases of this research.)

ARR computes the Value, or "expected utility," of a route ~~is computed~~ by the following formula, based on Bayesian decision theory:

> Value of Route = (the probability of arriving at and damaging the target) x (the value of the target) - (the probability of the aircraft being destroyed anywhere on the route) x (the value of the aircraft).

This equation highlights two important features of route planning or replanning: (1) two major uncertainties must be considered: the probability of damaging the target and the probability of own aircraft destruction (lethality); hence, it distinguishes between risks on the ingress and risks associated with the entire route; and (2) it requires a comparison between target value and the value of friendly aircraft. In essence, what this equation says is that for a route to be acceptable, the chance of damaging the target (i.e., success on ingress) and the value of the target must be great enough to outweigh the chance of being destroyed.

Tradeoffs involving these factors may be critical in route replanning when a pop-up threat appears during the ingress. For example, two revised route options for avoiding a pop-up threat may be available, which differ in how they allocate risk between ingress and egress. Route A plays it safe on the ingress, detouring significantly to avoid the pop-up threat; but on egress it must pass quite close to another threat due to fuel constraints. Route B takes a more direct path to the target than Route A, placing it in jeopardy from the pop-up threat, but leaving it with enough fuel on egress to avoid the other threat. It might be that Route B is on the whole safer (i.e., has a lower total lethality); but Route A might be preferable, even so, because it affords a better chance at the target. According to this model, choice between Route A and Route B depends on *how much* chance of damaging the target is worth *how much* risk to own aircraft. The present system takes such tradeoffs into account (through the above equation) in its evaluation of routes.

The Bayesian approach just described ignores the fact that inferential arguments underlying the system's evaluation of candidate routes may involve varying degrees of unreliability. As a result, the evidence may not uniquely determine an evaluation "score" for each route according to the above formula. Nevertheless, in a successful rerouting aid the potential lethality of a route to own aircraft and the likely damage inflicted on the target must be summarized in some way; different routes must be compared; and recommendations must be made to the pilot in a timely fashion.

ARR extends the traditional decision-theoretic approach to accommodate these requirements. It provides two lethality measures for each route: a lower lethality measure representing the lowest danger consistent with the evidence,

and an upper lethality measure representing the greatest danger consistent with the evidence. These measures are computed by appropriately reallocating uncommitted support, i.e., support assigned to subsets of hypotheses, to the elementary hypotheses in those subsets. Similarly, ARR provides both a lower and an upper measure of the chance of arriving at and damaging the target. By this means, the system computes an upper and a lower Value for each route: The upper Value is obtained by utilizing the lower measure for lethality and the upper measure for damaging the target; hence, it represents the most optimistic assessment of the route that is consistent with present evidence. The lower value is obtained by utilizing the upper measure for lethality and the lower measure for damaging the target; hence, it represents the most pessimistic supportable assessment.

Determination of a route recommendation (from a set of previously generated routes) now proceeds in two stages: (1) If the lower Value measure of a route is higher than the upper Value measure for another route, the first route is clearly preferred (and is recommended by our system). (2) In other cases, however, where the Value intervals for different routes overlap, the *evidence* available to the system is insufficient for a definitive choice (assuming that all cost-effective information collection options have been exhausted). In these cases, the system utilizes one of two normatively defensible, user-selected "decision attitudes" to determine a route recommendation. According to the *pessimism* (or worst case) attitude, upper measures are used for the bad outcome (i.e., lethality) and lower measures are used for the good outcome (i.e., damaging the target). According to the *conservatism* attitude, Values are computed utilizing the above equation with lower measures for all outcomes (i.e., uncommitted support is disregarded rather than reallocated, and only what the evidence positively supports is considered).

The following sections describe the inference mechanism in more detail, and present sample products of its operation.

4.2 The Belief Functions

ARR begins with prior area intelligence about threats in an area A (for simplicity, we assume A to be 2-dimensional space). This intelligence is represented as a belief function on A. En route, the system is notified of new

evidence of a threat in A, again expressed as a belief function on A. ARR must make inferences about whether the second item of evidence represents a new threat or the same threat, and if the same threat, whether it has moved to a new location.

To make this inference, ARR first extends the two belief functions to the set A x A x T, where the two copies of A represent the system's knowledge about the two threat localizations, and T = {S,D} is the set indicating whether the two signals represent the same or different threats. The elements of A x A x T are interpreted as follows.

(x,x,S) :	Same threat, unchanged location x.
(x,y,S) :	Original threat at x has moved to location y.
(x,y,D) :	Different threats at locations x and y.

Now ARR must incorporate its prior knowledge about whether threats are likely to move, and if so, how far; as well as its knowledge about whether there are likely to be gaps in area intelligence, so that some threats may have been missed; and its knowledge about the typical or expected spacing of separate threats. Each of these items of evidence can also be expressed as a belief function over A x A x T.

ARR's evidence is summarized by five belief functions, described below. Table 4-1 defines formal notation for the focal elements of each of the belief functions and the belief assigned to each. A finite number of focal elements is assumed for each belief function.

Bel_1: Summarizes prior evidence about the location of the first threat. Belief is focused on circles of increasing radius centered at a_1. This evidence provides no information about the location of the second (or moved) threat or about whether the two threats are the same or different.

Bel_2: Summarizes evidence about the location of the second threat. Belief is focused on circles of increasing radius centered at a_2. This evidence provides no information about the original threat or about whether the two threats are the same or different.

Bel_3: Summarizes evidence about movement. This evidence provides no information about whether the two threats are the same or different, but if they are the same, there is evidence about whether there was movement (e.g., observed transport activity) and if so, how much (based on time available and estimated speed capabilities). Thus, belief is focused on the diagonal H in AxA (threats in the

Belief Function	Focal Element	Belief Assignment	
Bel_1	$S_x(a_1) \times A \times \{S,D\}$	$m_1(x)$	$(\Sigma m_1(x)=1)$
Bel_2	$A \times S_y(a_2) \times \{S,D\}$	$m_2(y)$	$(\Sigma m_2(y)=1)$
Bel_3	$(H \times \{S\}) \cup (A \times A \times \{D\})$ $(C_w \times \{S\}) \cup (A \times A \times \{D\})$	s $m_3(w)$	 $(\Sigma m_3(w)=1-s)$
Bel_4	$A \times A \{S\}$ $A \times A \times \{S,D\}$	q $1-q$	
Bel_5	$(B_z \times \{D\}) \cup (A \times A \times \{S\})$	$m_5(z)$	$(\Sigma m_5(z)=1)$

Definition of symbols:

$S_x(a_1) = \{a : |a-a_1| \leq x\}$

$S_y(a_2) = \{a : |a-a_2| \leq y\}$

$H = \{(a,a) : a \varepsilon A\}$

$C_w = \{(a,b) : f_*(w) \leq |a-b| \leq f^*(w)\}$

(where $f_*(w)$ and $f^*(w)$ are lower and upper bounds for the distance range of the set C_w)

$B_z = \{(a,b) : g_*(z) \leq |a-b| \leq g^*(z)$

(where $g_*(z)$ and $g^*(z)$ are lower and upper bounds for the distance range of the set B_z)

Table 4-1: Summary of Belief Functions for Pilot Aid

same location), and on sets C_w, each of these representing a range of distances the threat might have moved.

Bel_4: Summarizes evidence about the thoroughness of intelligence. Belief is focused on subsets of T (whether the threats are the same or different); this evidence provides no evidence about location or separation of the threats.

Bel_5: Summarizes evidence about the separation of different threats (if threats are different, they are likely to be separated). Belief is focused on sets B_z, each of these representing a range of distances the threats might be separated. This belief function provides no information about whether the threats are the same or different, or about their separation in case they are the same.

These five belief functions may be combined by Dempster's Rule to obtain a belief function Bel_* over A x A x T. Table 4-2 summarizes the focal elements of the combined belief function and the belief assigned to each. The belief function Bel_* is obtained by normalizing these belief assignments (dividing by 1 minus the total belief assigned to the null set). The steps in combining the belief functions are summarized below.

1. Combine Bel_1 and Bel_2 to obtain a new belief function with focal elements $S_x(a_1) \times S_y(a_2) \times \{S,D\}$, each with belief $m_1(x)m_2(y)$.

2. Combine Bel_3 and Bel_5 to obtain a new belief function with focal elements $(H \times \{S\}) \sqcup (B_z \times \{D\})$ (belief $sm_5(z)$) and $(C_w \times \{S\}) \sqcup (B_z \times \{D\})$ (belief $m_3(w)m_5(z)$). The first type of focal element represents the belief that if the threats are the same, there is no movement; otherwise, their separation is described by the range of distances represented by B_z. The second type of focal element represents the belief that if there is movement, it is represented by C_w; if the threats are different, separation is described by B_z.

3. Combining the results of Steps 1 and 2 gives a belief function with two types of focal element. The first type of focal element is $[(S_x(a_1) \times S_y(a_2)) \cap H \times \{S\}]$ $[(S_x(a_1) \times S_y(a_2)) \cap B_z \times \{D\}]$ representing belief in a single threat unmoved, or a different threat. The second type of focal element is $[(S_x(a_1) \times S_y(a_2)) \cap C_w \times \{S\}]$ $[(S_x(a_1) \times S_y(a_2)) \cap B_z \times \{D\}]$, representing belief in a moved or a different threat.

4. The final step is to add in Bel_4, the belief in a single threat. The result is the same focal elements as in 3 (representing the uncommitted belief), as well as the focal elements $S_x(a_1) \times S_y(a_2) \cap H \times \{S\}$ and $S_x(a_1) \times S_y(a_2) \cap C_w \times \{S\}$, representing belief in a single threat with unchanged location or with location change described by C_w.

Table 4-2a Combined Belief Assignments

Focal Element	Belief (non-normalized)	Condition	Belief Assignment (U=unchanged, M=moved, D=differen
① $[S_x(a_1) \times S_y(a_2) \cap H] \times \{S\}$ $\cup [S_x(a_1) \times S_y(a_2) \cap B_z] \times \{D\}$	$(1-q)sm_1(x)m_2(y)m_5(z)$	$S_x(a_1) \cap S_y(a_2) \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z \neq \emptyset$	U,D
		$S_x(a_1) \cap S_y(a_2) = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z \neq \emptyset$	D
		$S_x(a_1) \cap S_y(a_2) \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z = \emptyset$	U
		$S_x(a_1) \cap S_y(a_2) = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z = \emptyset$	$\emptyset$
② $[S_x(a_1) \times S_y(a_2) \cap H] \times \{S\}$	$qsm_1(x)m_2(y)$	$S_x(a_1) \cap S_y(a_2) \neq \emptyset$	U
		$S_x(a_1) \cap S_y(a_2) = \emptyset$	$\emptyset$
③ $[S_x(a_1) \times S_y(a_2) \cap C_w] \times \{S\}$ $\cup [S_x(a_1) \times S_y(a_2) \cap B_z] \times \{D\}$	$(1-q)m_1(x)m_2(y)m_3(w)m_5(z)$	$S_x(a_1) \times S_y(a_2) \cap C_w \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z \neq \emptyset$	M,D
$(H \not\subset C_w)$		$S_x(a_1) \times S_y(a_2) \cap C_w = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z \neq \emptyset$	D
		$S_x(a_1) \times S_y(a_2) \cap C_w \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z = \emptyset$	M
		$S_x(a_1) \times S_y(a_2) \cap C_w = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z = \emptyset$	$\emptyset$

Table 4-2b Combined Belief Assignments

Focal Element	Belief (non-normalized)	Condition	Belief Assignment (U=unchanged, M=moved, D=differen
4 $[S_x(a_1) \times S_y(a_2) \cap C_w] \times \{S\} \cup [S_x(a_1) \times S_y(a_2) \cap B_z] \times \{D\}$	$(1-q)m_1(x)m_2(y)m_3(w)m_5(z)$	$S_x(a_1) \cap S_y(a_2) \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z \neq \emptyset$	U,M,D
(same as 3, but $H \subset C_w$)		$S_x(a_1) \cap S_y(a_2) = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap C_w \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z \neq \emptyset$	M,D
		$S_x(a_1) \times S_y(a_2) \cap C_w = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z \neq \emptyset$	D
		$S_x(a_1) \cap S_y(a_2) \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z = \emptyset$	U,M
		$S_x(a_1) \cap S_y(a_2) = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap C_w \neq \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z = \emptyset$	M
		$S_x(a_1) \times S_y(a_2) \cap C_w = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap B_z = \emptyset$	$\emptyset$
5 $[S_x(a_1) \times S_y(a_2) \cap C_w] \times \{S\}$	$qm_1(x)m_2(y)m_3(w)$	$S_x(a_1) \times S_y(a_2) \cap C_w \neq \emptyset$	M
($H \not\subset C_w$)		$S_x(a_1) \times S_y(a_2) \cap C_w = \emptyset$	$\emptyset$

Table 4-2c: Combined Belief Assignments

Focal Element	Belief (non-normalized)	Condition	Belief Assignment (U=unchanged, M=moved, D=differe
6 $[S_x(a_1) \times S_y(a_2) \cap C_w] \times \{S\}$ (same as 5, but $H \subseteq C_w$)	$qm_1(x)m_2(y)m_3(w)$	$S_x(a_1) \cap S_y(a_2) \neq \emptyset$	U,M
		$S_x(a_1) \cap S_y(a_2) = \emptyset$ $S_x(a_1) \times S_y(a_2) \cap C_w \neq \emptyset$	M
		$S_x(a_1) \times S_y(a_2) \cap C_w = \emptyset$	$\emptyset$

A belief function over the three possibilities unchanged (U), moved (M), and different (D) can be defined by the marginal belief function of Bel_*. Table 4-3 gives this marginal belief function.

Table 4-1 presents quite complex mathematical conditions for belief in the three hypotheses (unchanged, moved, different) and their various combinations. These conditions can be greatly illuminated by describing them verbally in terms of *reasons* for belief in the hypotheses.

- *Contours: degree of overlap* - The threat can be *unchanged* only if it is possible that the two signals could have come from the same location. Thus, belief in an unchanged threat is supported to the degree that the location contours of Bel_1 and Bel_2 overlap.
- *Distance: possibility of movement* - *Movement* is possible only when the separation between the threats (described by Bel_1 and Bel_2) is consistent with the distance a threat might have moved (as described by the distance ranges of Bel_3). Thus, belief in a moved threat is supported to the extent that the contours of Bel_1 and Bel_2 are consistent with the distance ranges of Bel_3.
- *Likelihood of movement* - Bel_3 also has a focal element representing positive belief in no movement. Belief in *movement* is supported to the extent that this focal element has low belief.
- *Distance: possibility for different threats* - *Different threats* are possible only when the threat separation (described by Bel_1 and Bel_2) is consistent with threat spacing information (described by Bel_5). Thus, belief in different threats is supported to the extent that the contours of Bel_1 and Bel_2 are consistent with the distance ranges of Bel_5.
- *Threat coverage* - The threats can be *different* only if it is possible that a threat was missed. Thus, belief in different threats is supported by a low belief in the thoroughness of intelligence (as described by Bel_4).

The system has the capability for selecting which of the above reasons provides primary support either for or against each of the hypotheses. The display informs the user, then, not only of the belief in each hypothesis, but of the reasons supporting that belief.

Table 4-3: Marginal Belief Function over {U,M,D}

Focal Element	Belief
{U}	$Bel_{*}(Hx\{S\})$
{M}	$Bel_{*}(Hx\{S\})$
{D}	$Bel_{*}(AxAx\{D\})$
{U,M}	$Bel_{*}(AxAx\{S\})$
{U,D}	$Bel_{*}(AxAx\{D\}) + Bel_{*}(Hx\{S\})$
{M,D}	$Bel_{*}(AxAx\{D\}) + Bel_{*}(Hx\{S\})$
{U,M,D}	1

4.3 Conflict Resolution

We see from Table 4-1 that combining belief functions with mass focused on incompatible subsets results in belief assigned to the null set. We use as a measure of conflict the amount of belief assigned to the null set when applying Dempster's Rule. This is the measure of conflict used by Shafer (1976) and by Cohen (1985).

Viewed in another way, conflict occurs to the extent that there is evidence *against* all three hypotheses (unchanged, moved, different). For each of the six null-set entries in Table 4-1, we can identify the reasons for the occurrence of the conflict, in terms of the taxonomy given at the end of the previous section. Table 4-4 gives the reasons for the six types of conflict. For example, the first type of conflict occurs to the extent that belief is assigned to non-overlapping contours, to non-movement, and to distance ranges incompatible with location contours.

When combination of the five belief functions results in conflict greater than threshold t_c, the system's conflict resolution procedure is invoked. The conflict resolution procedure is, in effect, a mechanism for searching within the arguments leading to each of the five belief functions in an attempt to identify potential weaknesses in the arguments. When information about such weaknesses ~~are~~ is obtained ~~identified~~, the corresponding belief functions are discounted, leading to reduction in conflict. As long as potential weaknesses can be identified, the process of conflict resolution continues until conflict is reduced to below t_c.

CONFLICT TYPE (From Table 4-6)	CONTOUR NON-OVERLAP	DISTANCE PRECLUDES MOVEMENT	EVIDENCE FOR MOVE-MENT (if same threat)	EVIDENCE AGAINST MOVEMENT	DISTANCE PRECLUDES DIFFERENT	COVERAGE GOOD
(1)	X			X	X	
(2)	X			X		X
(3)		X	X		X	
(4)	X	X			X	
(5)		X	X			X
(6)	X	X				X

Table 4-4: Reasons for Conflict--Six Types of Conflict

As indicated in Figure 3-1, each belief function has associated with it a set of potential *discount factors*, or factors influencing the reliability of the link between the evidence and the conclusion. An example of such a discount factor would be the presence of ECM in the area when a SAR signal is observed; such presence would tend to discredit the location estimate derived from the SAR signal.

Each discount factor has associated with it an *initial* belief function. This function represents the "default" assumption the system wishes to make about the presence of the factor, prior to testing for its presence. A reasonable initial belief function might be vacuous, assigning all belief to the universal set, and thus representing no information about the factor's presence or absence. Alternatively, there might be evidence available initially, whether specific to this mission, or based on prior experience. The initial belief function allows such information to be incorporated into the analysis.

Each discount factor also has an associated *test* for factor presence. The test may have several possible outcomes. Each test outcome o is associated with a belief function which summarizes the impact of observing test outcome o on belief in factor presence. If a test is performed with outcome o, the associated belief function is combined via Dempster's Rule with the initial belief function to obtain an updated belief function for factor presence.

Belief functions are discounted according to the amount of belief directly committed to factor *presence*. (This practice corresponds to an assumption that the evidential link is valid until evidence is observed to the contrary.) If B_k is the amount of belief directly committed to the presence of discount factor k, then the discount rate for the corresponding belief function is:

1 −

$$\delta = \sum_k w_k B_k$$

where the summation is over those discount factors associated with the given belief function. The number w_k is a measure of the impact of the presence of factor k on the discounting. The w_k are assumed to be positive and to sum to 1. The resulting discount rate ranges between 0 and 1, with a discount rate of 0 corresponding to complete discrediting of the evidential link, and a discount rate of 1 corresponding to no discounting relative to the initial belief function.

The belief function Bel_i is discounted by multiplying the belief associated with each focal element by the discount rate δ. This results in beliefs summing to less than 1; belief is now added to the universal set to correspond to the belief subtracted from each of the focal elements.

When the initial belief functions for some of the discount factors for Bel_i are non-vacuous, the system must discount belief function Bel_i before initial application of Dempster's Rule (described in Section 2.2.2). Thus, the initial pass of the inference mechanism incorporates any discounting deemed appropriate prior to combination of the evidence.

We now describe the process initiated when combination of initial belief functions (which incorporate the initial discount rates) results in conflict exceeding the threshold t_c. ARR moves to the "second pass" of its inference mechanism, described by the following five steps.

1. Decide which discount factor for which belief function is the provisional "culprit" and which test to perform on the culprit. (This step is the crux of the algorithm and the selection criteria are discussed in detail below.) If no culprit can be found, initiate Pass 3.

2. Perform the test and revise belief in the appropriate discounting factor.

3. Compute a revised discount rate and apply it to the culprit belief function, resulting in a new belief function.

4. Recombine the belief functions according to Dempster's Rule, as described in Section 2.2.2. The result is a new combined belief function, and a new measure of conflict.

5. If conflict is below t_c, stop. Otherwise, return to Step 1.

The test chosen is based on potential for conflict reduction, balanced against the cost of performing the test.

Each test, then, must have a cost associated with it. It is in these costs that a crucial difference between ground-based and in-flight aids comes to the fore. An in-flight aid would associate a very high cost with performing any test for which results could not be obtained very quickly. Moreover, the costs might change dynamically with flight progress (some tests being feasible early on when the time stress is not so great, but becoming infeasible later in the mission).

To decide which test to perform, our prototype system first evaluates each test to see which has the maximum potential for conflict reduction. To do this, the system computes a measure of the impact of discounting each of the component belief functions on conflict (a partial derivative of conflict with respect to the discount rate on the component belief function). It then identifies for each test associated with each belief function the maximum potential for discounting the associated belief function (taken over all test results). This maximum discount rate is multiplied by the partial derivative to obtain a measure of the maximum possible impact on conflict for the given test. This quantity is then divided by cost to obtain a measure of maximum conflict reduction per unit cost. The test is chosen for which this measure is the highest. (Note that there is no guarantee that conflict reduction will be as great as indicated by this measure--the measure is based on the *most favorable* result for the test, which may not be the result observed.)

Formally, let δ_i be the initial discount rate for belief function i, and let c be the conflict computed under discount rate δ_i (i.e., the conflict from Pass

1 of the algorithm). Let c_i be the partial derivative of c with respect to δ_i. Now, for a given test t, let $\delta^*(t)$ be the maximum discount rate (over all test results) that can be obtained for that test. Now, for each discount factor f, let t(f) be the test for factor presence, $\gamma(t(f))$ the cost of the test, and i(f) the index of the associated belief function. Then let

$$u(f) = -(\delta^*(t(f))-\delta_i)c_i/\gamma(t(f));$$

the quantity u(f) can be thought of as the utility of testing for discount factor f (the negative sign occurs because conflict varies inversely with discount rate).

The test t(f) is performed for the factor f for which u(f) is maximized.

Even in the face of extreme conflict, discounting a contributing argument may not be appropriate. There must be a reasonably strong case that the most likely cause (or causes) of the conflict have been identified. It may be the case that the system cannot find a test to perform, whether because no test has the potential for significant conflict reduction, because all tests cost too much to perform, or because all possible tests have already been performed. We saw above that in this case, the system resorts to a third pass of discounting all belief functions, according to a formula by which those belief functions contributing most to the conflict are discounted the most. The mechanism for this across-the-board discounting is quite a natural extension of the present framework for representing evidential reasoning. Each evidential argument is associated with a discount factor called "conflict with other evidence." Across-the-board discounting involves an increase in the belief in the presence of this factor, proportional to the contribution of a given argument to the conflict. The weight on this factor, for a given argument, reflects the firmness with which the system will retain commitment to that particular evidential link in the face of conflicting data.

4.4 Sample Results

In this section, we describe some of the results produced by applying the inference mechanisms within our prototype system to sample data. We stress that the following discussion is not meant as a description of the user-system

interface; these results would obviously not all be presented to users in this form (see Section 4.5 below).

Figures 4-1a through 4-1e illustrate the output of Pass 1 through the system's inference mechanism, for each of five different inputs. Each of these figures describes the input belief functions, the combined threat classification belief function, and the amount of conflict in the evidence.

In the first analysis (Figure 4-1a), belief contours for the first and second threats are centered at a distance of 5 units (say, miles) apart. The location contours describe how certain we are of these localizations. Thus, for the first localization, there is belief of .18 that the threat lies within .93 units of the center (2,2). There is belief .54 (.18 + .18 + .18) that the threat lies within 1.5 units of (2,2). We see that the localization of the second threat is less precise than that of the first--the belief contours have greater radius. Belief of .3 is committed directly to the hypothesis that the second threat, if the same, has not moved; belief of .10 is uncommitted about whether or how far it moved, and the rest of the belief is distributed across nested intervals of distances the threat may have moved. Belief of .7 has been assigned to the second localization representing the same threat as the first (indicating fairly high confidence in area intelligence). Finally, if the threat has moved, we place belief .17 on its having moved at least 6.1 miles, and belief .68 (.17 + .17 + .17 + .17) on its having moved at least 3.2 miles.

The resulting belief function places the highest weight on the hypothesis of a single threat that has moved. This result is consistent with our confidence in area intelligence, as well as a small amount of belief placed on the threats being the same. Conflict is not too large, at a level of .17. (We have tentatively adopted a conflict threshold of .25 for initiation of Pass 2; more experience is needed to determine what level is best in this application.)

The second set of inputs (Figure 4-1b) is the same as the first, except that the belief assigned to an unmoved threat has been raised from .3 to .7, with corresponding reductions in belief for the intervals the threat might have moved. The result is what we would expect: relative to the first set of

Input Belief Functions:

Bel_1: Center of Contours = (2,2)

Radius	.44	.93	1.5	2.2	3.3	5.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_2: Center of Contours = (5,6)

Radius	.88	1.9	3.9	4.3	6.6	10.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_3: Belief Assigned to Diagonal = 0.3

Lower Distance	1.02	.88	.75	.63	0.0
Upper Distance	1.34	1.53	1.78	2.14	∞
Committed Belief	.15	.15	.15	.15	.10

Bel_4: Belief Assigned to Same Threat = 0.7

Bel_5:

Lower Distance	6.1	4.9	4.0	3.2	2.6	2.0
Upper Distance	∞	∞	∞	∞	∞	∞
Committed Belief	.17	.17	.17	.17	.17	.15

Combined Belief Function: Classification of Second Threat
(U = unchanged; M = moved; D = different)

$Bel_*(\{U\}) = .15$	$Bel_*(\{U,M\}) = .63$
$Bel_*(\{M\}) = .43$	$Bel_*(\{U,D\}) = .31$
$Bel_*(\{D\}) = .09$	$Bel_*(\{M,D\}) = .71$
	$Bel_*(\{U,M,D\}) = 1.0$

Conflict (Mass Assigned to Null Set) in Combined Belief Function = .17

Figure 4-1a: Output of Pass 1 of Inference Mechanism

Input Belief Functions:

Bel_1: Center of Contours = (2,2)

Radius	.44	.93	1.5	2.2	3.3	5.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_2: Center of Contours = (5,6)

Radius	.88	1.9	3.9	4.3	6.6	10.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_3: Belief Assigned to Diagonal = 0.7

Lower Distance	1.02	.88	.75	.63	0.0
Upper Distance	1.34	1.53	1.78	2.14	∞
Committed Belief	.05	.05	.05	.05	.10

Bel_4: Belief Assigned to Same Threat = 0.7

Bel_5:

Lower Distance	6.1	4.9	4.0	3.2	2.6	2.0
Upper Distance	∞	∞	∞	∞	∞	∞
Committed Belief	.17	.17	.17	.17	.17	.15

Combined Belief Function: Classification of Second Threat
(U = unchanged; M = moved; D = different)

$Bel_*(\{U\}) = .38$
$Bel_*(\{M\}) = .18$
$Bel_*(\{D\}) = .12$

$Bel_*(\{U,M\}) = .61$
$Bel_*(\{U,D\}) = .66$
$Bel_*(\{M,D\}) = .38$
$Bel_*(\{U,M,D\}) = 1.0$

Conflict (Mass Assigned to Null Set) in Combined Belief Function = .22

Figure 4-1b: Output of Pass 1 of Inference Mechanism

Input Belief Functions:

Bel_1: Center of Contours = (2,2)

Radius	.44	.93	1.5	2.2	3.3	5.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_2: Center of Contours = (3,4)

Radius	.88	1.9	3.9	4.3	6.6	10.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_3: Belief Assigned to Diagonal = 0.7

Lower Distance	1.02	.88	.75	.63	0.0
Upper Distance	1.34	1.53	1.78	2.14	∞
Committed Belief	.05	.05	.05	.05	.10

Bel_4: Belief Assigned to Same Threat = 0.7

Bel_5:

Lower Distance	6.1	4.9	4.0	3.2	2.6	2.0
Upper Distance	∞	∞	∞	∞	∞	∞
Committed Belief	.17	.17	.17	.17	.17	.15

Combined Belief Function: Classification of Second Threat
(U = unchanged; M = moved; D = different)

$Bel_*(\{U\}) = .48$ $\quad$ $Bel_*(\{U,M\}) = .70$
$Bel_*(\{M\}) = .15$ $\quad$ $Bel_*(\{U,D\}) = .69$
$Bel_*(\{D\}) = .01$ $\quad$ $Bel_*(\{M,D\}) = .22$
$Bel_*(\{U,M,D\}) = 1.0$

Conflict (Mass Assigned to Null Set) in Combined Belief Function = .04

Figure 4-1c: Output of Pass 1 of Inference Mechanism

Input Belief Functions:

Bel_1: Center of Contours = (2,2)

Radius	.44	.93	1.5	2.2	3.3	5.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_2: Center of Contours = (8,8)

Radius	.88	1.9	3.9	4.3	6.6	10.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_3: Belief Assigned to Diagonal = 0.3

Lower Distance	1.02	.88	.75	.63	0.0
Upper Distance	1.34	1.53	1.78	2.14	∞
Committed Belief	.15	.15	.15	.15	.10

Bel_4: Belief Assigned to Same Threat = 0.3

Bel_5:

Lower Distance	6.1	4.9	4.0	3.2	2.6	2.0
Upper Distance	∞	∞	∞	∞	∞	∞
Committed Belief	.17	.17	.17	.17	.17	.15

Combined Belief Function: Classification of Second Threat

(U = unchanged; M = moved; D = different)

$Bel_*(\{U\}) = .02$ $\quad$ $Bel_*(\{U,M\}) = .14$

$Bel_*(\{M\}) = .11$ $\quad$ $Bel_*(\{U,D\}) = .59$

$Bel_*(\{D\}) = .51$ $\quad$ $Bel_*(\{M,D\}) = .89$

$Bel_*(\{U,M,D\}) = 1.0$

Conflict (Mass Assigned to Null Set) in Combined Belief Function = .18

Figure 4-18: Output of Pass 1 of Inference Mechanism

Input Belief Functions:

Bel_1: Center of Contours = (2,2)

Radius	.44	.93	1.5	2.2	3.3	5.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_2: Center of Contours = (8,8)

Radius	.88	1.9	3.9	4.3	6.6	10.0
Committed Belief	.18	.18	.18	.18	.18	.10

Bel_3: Belief Assigned to Diagonal = 0.3

Lower Distance	1.02	.88	.75	.63	0.0
Upper Distance	1.34	1.53	1.78	2.14	∞
Committed Belief	.15	.15	.15	.15	.10

Bel_4: Belief Assigned to Same Threat = 0.7

Bel_5:

Lower Distance	6.1	4.9	4.0	3.2	2.6	2.0
Upper Distance	∞	∞	∞	∞	∞	∞
Committed Belief	.17	.17	.17	.17	.17	.15

Combined Belief Function: Classification of Second Threat

(U = unchanged; M = moved; D = different)

$Bel_*(\{U\}) = .08$
$Bel_*(\{M\}) = .38$
$Bel_*(\{D\}) = .31$

$Bel_*(\{U,M\}) = .49$
$Bel_*(\{U,D\}) = .42$
$Bel_*(\{M,D\}) = .85$
$Bel_*(\{U,M,D\}) = 1.0$

Conflict (Mass Assigned to Null Set) in Combined Belief Function = .42

Figure 4-1e: Output of Pass 1 of Inference Mechanism

inputs, belief assigned to an unchanged threat has increased, and belief in a moved threat has decreased. Conflict has increased a little, and is now near the threshold for initiation of Pass 2. This reflects the fact that the overlap in the location contours is little enough that there is reasonable conflict in attributing both to an unmoved threat.

Now consider a third set of inputs identical to the second, except that the centers of the contours move closer together (Figure 4-1c). As expected, belief in an unchanged threat is greatly increased relative to that in a moved threat. Moreover, conflict has decreased to nearly zero, indicating the extent to which the conflict in the second set of belief functions was due to non-overlapping contours.

The fourth set of inputs (Figure 4-1d) is the same as the first, except that the centers of the location contours are now farther apart, and belief in the thoroughness of area intelligence has decreased to 0.3 (indicating a fairly high possibility that a threat may have been overlooked). The result is a high belief in the two localizations representing different threats. The conflict level, .18, is not sufficiently high to trigger conflict reduction.

Our final example (Figure 4-1e) illustrates what happens when the threat localizations remain widely separated, but confidence in area intelligence is raised again to .7. The combined belief functions assign nearly equal weight to moved and different threats (the overlap in the contours being small enough that very little belief is assigned to an unchanged threat). But most importantly, conflict is greatly increased; nearly half of the belief in the combined functions is assigned to the null set. This set of inputs results in triggering of Pass 2, the conflict reduction step.

In Pass 2, our system first chooses to test for the presence of ECM in the area. The result is a discounting of the belief contours of Bel_2 by a discount rate of .31. After discounting, Dempster's Rule is reapplied (Figure 4-2), and the new level of conflict is greatly reduced to 0.29. This level remains above the threshold of .25, so a second pass of conflict reduction is initiated. The system next chooses to reassess the thoroughness of area intelligence. It searches for information about area intelligence (whether from the pilot or by querying ground-based sources), and decides to discount

Bel_4 by a discount rate of .19. After this final discounting, conflict is reduced to an acceptable .23. Figure 4-2 reports the system's inferences about threat classification. Most of the mass has been allocated to a combination of the hypotheses that the threat has moved or is different (total mass .76 in the final pass).

Relative to Figure 4-1d, there is a higher degree of belief that the threat has moved. This is because the high confidence in area intelligence tends to discredit the hypothesis of different threats. Note that the final pass, in discounting Bel_4, has resulted in decreased confidence in area intelligence (.7 is discounted to $(1-.19) \times .7 = .57$). This results in higher belief in different threats after the final pass than on the previous pass.

In summary, we see that the numerical results in the examples given conform to intuition. Increasing initial belief in an unchanged threat increases conflict to the extent that contours do not overlap, and also increases final belief in an unchanged threat. Moving the signals closer together increases final belief in an unchanged threat; moving them far apart increases belief in different threats. Incompatible initial beliefs (threats far apart but missed threat unlikely) results in conflict, which is resolved by discounting.

4.5 Hardware

In the interests of efficiency of coding and portability, the demonstration system is implemented in the C language on an IBM AT with an 80287 coprocessor, at least 512KB of random access memory, and IBM Enhanced Graphics Adaptor (640 x 360 pixels with 16 simultaneous colors), and IBM Enhanced Graphics monitor, and a mouse input device. The mouse, though clearly inappropriate as a cockpit instrument, should provide an adequate functional simulation for demonstration purposes of other "pointing" input modes, such as eye movements. A fully function cockpit hardware configuration is likely to differ in other respects as required by the cockpit environment.

First Conflict Resolution Pass:

Classification of Second Threat

$Bel_*(\{U\}) = .14$	$Bel_*(\{U,M\}) = .58$
$Bel_*(\{M\}) = .40$	$Bel_*(\{U,D\}) = .37$
$Bel_*(\{D\}) = .17$	$Bel_*(\{M,D\}) = .74$
	$Bel_*(\{U,M,D\}) = 1.0$

Conflict in Assigned Belief Function = .29

Second Conflict Resolution Pass:

Classification of Second Threat

$Bel_*(\{U\}) = .10$	$Bel_*(\{U,M\}) = .43$
$Bel_*(\{M\}) = .30$	$Bel_*(\{U,D\}) = .41$
$Bel_*(\{D\}) = .23$	$Bel_*(\{M,D\}) = .76$
	$Bel_*(\{U,M,D\}) = 1.0$

Conflict in Assigned Belief Function = .23

Figure 4-2: Output of Pass 2 of Inference Mechanism

5.0 HUMAN COMPUTER INTERFACE FOR THE ADAPTIVE ROUTE REPLANNING AID

5.1 Basic Approach

The user computer interface for an effective inflight rerouting aid must successfully balance a set of competing objectives: it must (a) minimize demands on user time and effort, while at the same time (b) communicating both recommendations and *reasons* for those recommendations in a way which maximizes user understanding, and (c) permitting rapid, effective user inputs where they might be critical for mission success.

Traditional approaches to the human-computer interface have proven largely inadequate for achieving the multiple objectives outlined above. On the one hand, automated sensor and communication systems have amplified the volume of data available to users without providing significant assistance in the interpretation of that data and in its use for the decision-making process. On the other hand, expert systems and decision aids which have been more recently proposed have gone to the other extreme, by offering a single rigid approach to analysis and decision making. Few current systems have attempted to deal in a flexible manner with the diversity of decision-making situations in real-world combat environments and the variety of user-preferred problem-solving and decision-making styles. Our goal has been to sketch the design of an adaptive, highly flexible user interface to implement the inference mechanisms described in the previous sections. The goal is, ultimately, to produce an aid that is both *personalized* in the sense that it accommodates a variety of user-preferred knowledge representations and information-processing strategies, and *prescriptive*, in the sense that it encourages and in some cases prompts user actions that overcome deficiencies in the user-preferred approach.

At the highest level, the cognitive interface between a user and a computer-based decision aid, such as ARR, can be characterized by five generic functions (see Cohen, Thompson, and Chinnis, 1985):

- *Select*: Users may personalize displays of information by organizing them around alternative meaningful user-designated objects (e.g., time periods, spatial regions, options, components of options, at-

tributes of options). The user can examine any significant input, inference rule, intermediate conclusions, or final result concerning a given object.

- *Modify*: The user can alter values of any database element and immediately observe the impact on results downstream in a chain of reasoning; users may undo their modifications and restore the original values; user inputs may be at any level of fuzziness or precision.

- *Generate*: Users may define options at any level of abstractness, completeness, or precision, and with respect to any time horizon; automatic option generation procedures work within whatever constraints a user has provided.

- *Analyze*: In the evaluation of options, users may examine predicted outcomes according to any preferred scheme (e.g., static or temporal/dynamic; organizing information by attributes or by options), and may order the relative importance of different evaluative criteria to any degree of completeness/incompleteness and fuzziness or precision.

- *Alert*: The system prompts a user when events occur or facts are learned which would play a significant role in user-preferred modes of reasoning and organizing information.

Within the constraints of the present work, only a partial demonstration of the user interface has been implemented. The demonstration system consists of about 40 screens embedded within a "live" menu system. Although many of the screens reflect the output of the inference algorithm, others are "canned", and serve the purpose of illustrating the interface design in a fuller way than the inference mechanism implementation itself permits. The screens provide appropriate displays for a moderate number of menu requests, representing a specific route replanning example. Input and output operations for some of the live displays are also operational. A user who stays within the broad boundaries of the example may, therefore, get a fairly good feeling for the intended operation of the aid. The resulting demonstration system serves several important purposes as a design tool: by demonstrating the relevant AI inferencing technology in a specific subproblem; by providing a feel for the quality of the user-computer interface and an opportunity to validate its effectiveness; and by serving as an initial prototype to guide further development of the overall system.

5.2 Overview of the Interface

Interface features of the Adaptive Route Replanner are intended to minimize

the attention users must devote simply to operating the aid. Virtually all displays present information graphically, by a combination of maps and charts. All user inputs are by means of a single input device, which implements a pointing function. In the present demonstration this device is a mouse and associated function keys. In a final cockpit implementation however, the input device might involve touching a screen, eye movements, or any functionally equivalent method.

The Adaptive Route Replanner main menu includes the following items:

SITUATION RECOMMEND POSSIBILITIES REASONS CREDIBILITY LETHALITY ACCEPT

SITUATION provides the pilot a basic view of the current tactical environment (Figure 5-1). In the present demonstration, it displays threat danger contours which integrate all available information about threat locations, threat IDs, threat capabilities, and terrain. (In the demonstration system the derivation of these contours takes place in a "black box" and is represented by canned screens. Technically, such contours represent the differential change in the probability of destruction for an aircraft at the given location for a specified period of time. In a completed implementation, the user would have the additional capability of calling up displays which depict threat location uncertainty, ID uncertainty, capability, and terrain.) The situation display also shows the current route and the location of the aircraft on that route.

When a pop-up threat occurs, the situation display provides critical information (Figure 5-2). The display, however, minimizes the information processing burden on users by (a) alerting with regard to such a threat only when it *matters* (i.e., when the increase in lethality of the current route due to the new threat exceeds a preset threshold); (b) using displays that emphasize *changes* from the expected situation: i.e., areas where danger has increased due to the pop-up threat by x percent or more are highlighted in red; and (c) prompting users only in regard to those uncertainties in the evidence that are critical for the decision-making task (see below).

After the occurrence of a pop-up threat has been indicated by the aid, the user has several alternative courses of action available. (1) He may select

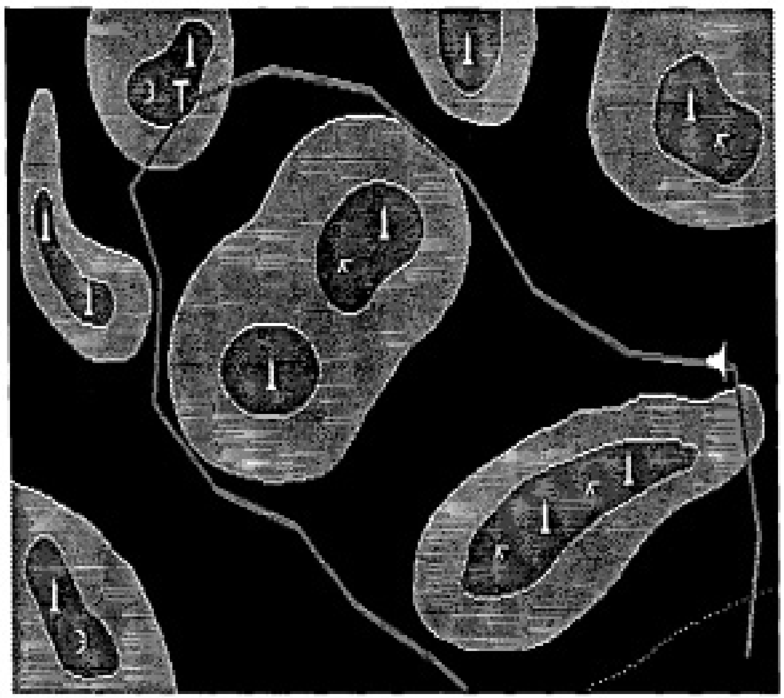


Figure 5-1

DATA: Radar signal at bearing = 183°, A/C at 38° long., 42° lat.

Source: SAR   Reason: New Threat

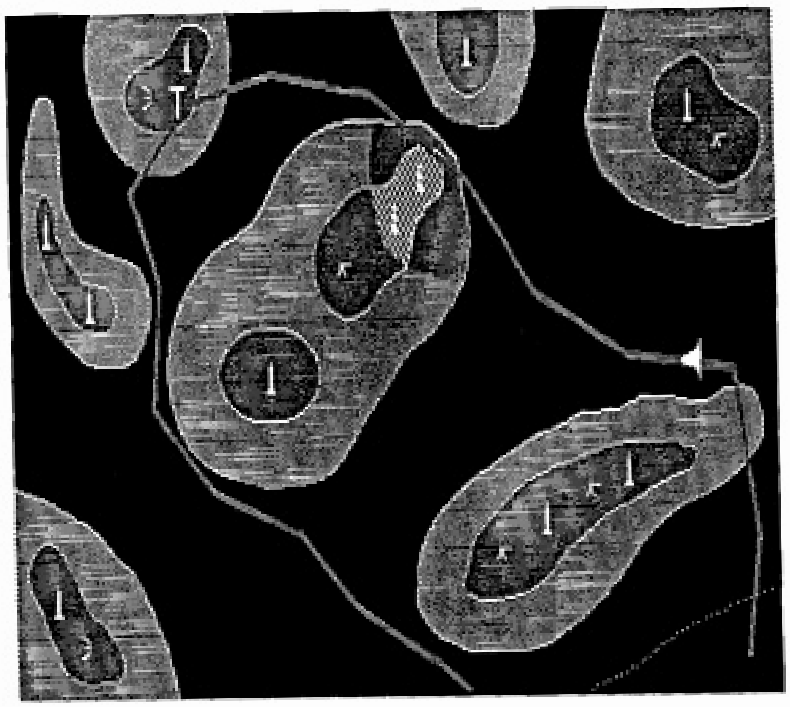


Figure 5-2

RECOMMEND, and the aid will automatically suggest a route which accommodates the new information about the pop-up threat (Figure 5-3); (2) he may participate in the route generation process by providing inputs which constrain the routes generated and recommended by the aid (this capability has not been demonstrated in the present system); (3) he may explore in greater detail the lines of reasoning underlying the SITUATION display and the route recommendation.

Option (3) is implemented by the REASONS and the CREDIBILITY displays. REASONS as shown in Figure 5-4 displays the main alternative hypotheses for interpreting the current data about the pop-up threat. In particular, in this example a SAR signal has been received by the aircraft and may represent (a) a new signal from a previously identified stationary threat, (b) a new signal from a previously identified threat which has moved, or (c) a signal from a previously unknown threat. The REASONS display shows the relative strength of each of these hypotheses given the available evidence, resulting from application of ARR's inference mechanism.

Perhaps more importantly, beneath each bar in this histogram is a "ledger" (P. Cohen) of *reasons* for or against that particular hypothesis. Thus, for example, support for the hypothesis that the SAR signal represents a stationary, previously identified threat might come from: the high overlap in location contours represented by the new signal and by the previous localization, intelligence assessments which indicate a low mobility for the threat type involved, or a high assessment of the thoroughness of prior area intelligence. Similarly, possible reasons in support of the hypothesis that the signal represents a moved threat include: relatively low overlap in the location contours for the two localizations, high confidence in the thoroughness of prior area intelligence, and an intelligence assessment that the relevant threat type does possess the requisite mobility. Finally, reasons that might support the possibility that a new threat has emerged in the area include: low confidence in the exhaustiveness of prior area intelligence and a distance between the two signals which conforms with our understanding of enemy siting practice. The REASONS display automatically indicates which of these possible reasons have *in fact* influenced the evaluation of the new signal.

DATA: Radar signal at bearing = 183°, A/C at 38° long., 42° lat.
Source: SAR   Reason: New Threat

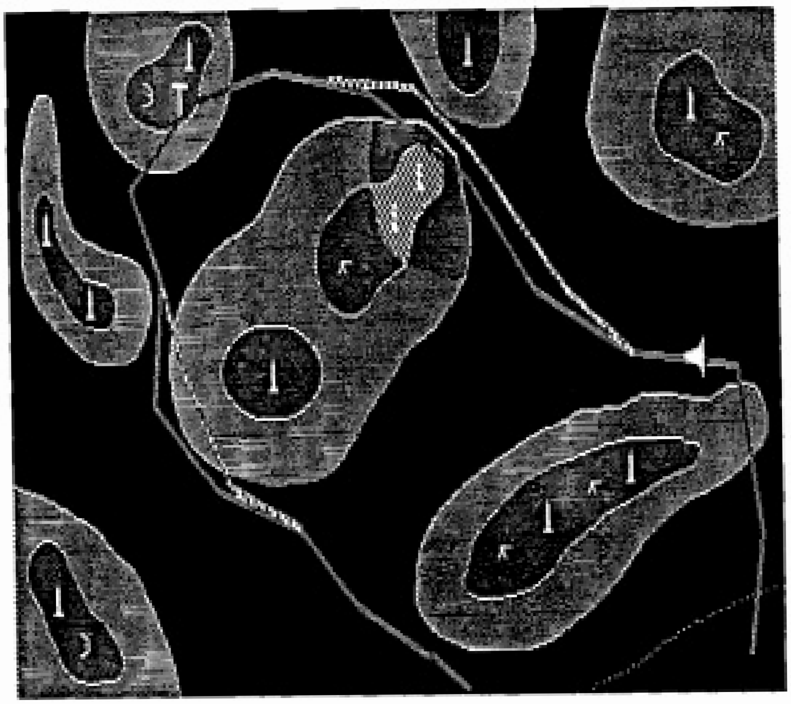


Figure 5-3

DATA: Radar signal at bearing = 183°, A/C at 38° long., 42° lat.

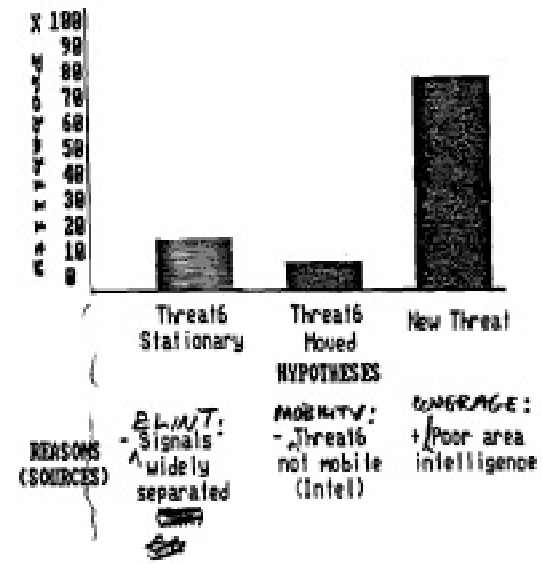


Figure 5-4

If he wishes, the user may explore in even greater depth the line of reasoning which le~~a~~d to the current situation assessment. He may do this by selecting the CREDIBILITY screen (Figure 5-5). CREDIBILITY displays a histogram which represents the degree of credibility or confidence in the sources of evidence underlying the REASONS screen. Thus in the present example evidence is available from a prior localization based on HUMINT, a new signal from SAR, and three different types of specific intelligence (concerning threat siting, mobility, and thoroughness of intelligence coverage).

In addition to indicating a degree of credibility for each of these sources, however, the CREDIBILITY screen shows the qualitative basis for such credibility judgments in a highly natural way, in the form of positive and negative "endorsements" (P.Cohen). Under each bar of the histogram is a set of credibility factors which influence, either positively or negatively, our confidence in the relevant source.

The Adaptive Route Planner allows the user to modify inputs or intermediate values at any level of analysis. Thus, the user may modify the degree of belief in the hypotheses represented in Figure 5-4, the degree of credibility assigned to different sources in Figure 5-5, or the status of credibility factors shown in the lower part of Figure 5-5. (Only the latter two have been implemented in the present demonstration.) In each case, the results of the adjustment are reflected in automatic inferences farther downstream in the reasoning.

For example, the pilot may feel greater (or lesser) confidence in the reliability of SAR evidence than does the automatic inference process. If so, he may adjust the credibility assessment simply by pointing to the desired height on the histogram. Alternatively, the user may have information regarding a specific credibility factor. In that case, he may adjust the status of the credibility factor (which in turn will have an automatic effect on the heights of the histogram). For example, if the aid indicates a high ground reflectance (based, for example, on DMA maps), but the pilot's direct observation indicates low reflectance, he may alter this factor by pointing to reflectance, cycling through a set of values (including low, moderate, high, and unknown), and selecting. Similarly he may adjust ECM (unlikely, possible, probable, and unknown), weather, and so on. As these credibility factor

~~DATA: Radar signal at bearing = 183°, A/C at 38° long, 42° lat.~~

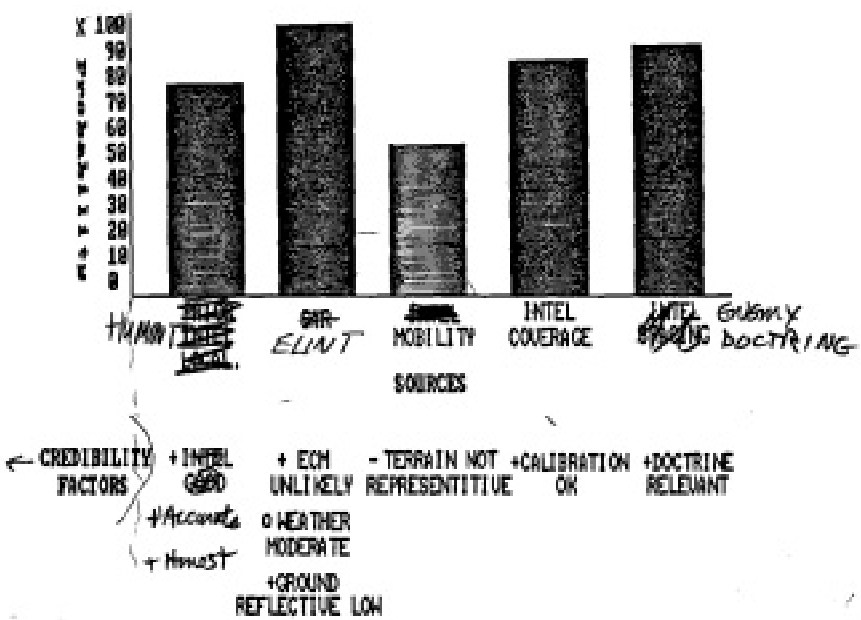


Figure 5-5

assessments are changed, the degree of support for relevant hypotheses in Figure 5-4 is adjusted automatically.

Up to this point, we have discussed the pilot's ability to explore (and to participate in) an inference task i.e., determining the nature and extent of the pop-up threat. But the pilot may also be concerned to evaluate the effectiveness of a recommended route revision. For this purpose, he may select the LETHALITY screen (Figure 5-6). This screen provides a representation of lethality, i.e., probability of own aircraft loss, as a function of time on a particular route. The two curves in Figure 5-6 represent the current route and the recommended route revision, respectively. Each curve is cumulative, showing how the risk on a route increases with time on that route. The final level on the ordinate for each route represents total chance of own aircraft destruction on that route. The slope on any given portion of a curve indicates the local danger in that portion of the route: steeper portions of the curve representing more dangerous areas and shallow portions of the curve representing less dangerous areas. I and T represent the initial point and the target, respectively, on each route. Generic symbols representing SAMs, radar, and artillery sites on these curves are keyed to corresponding symbols in the situation display.

At any time the user can indicate his acceptance of a route (either a recommended route or one generated by his own inputs) by selecting ACCEPT. When he does so, the accepted route becomes the new "current route" in all future displays.

It is inevitable that in many warfare situations, the available evidence will be incomplete or conflicting, or both. As discussed above, the present inference mechanism design is tailored to deal with those contingencies in a highly adaptive fashion. The interface design likewise in intended to facilitate user understanding and effective response to such contingencies. Figure 5-7 shows a SITUATION display in which significant inconsistency among sources of evidence has been indicated. The nature of the inconsistency is briefly summarized: the SAR signal may represent the presence of a new threat or it may originate from a previously identified stationary threat.

The inconsistency prompt occurs only when the system's automated processes of

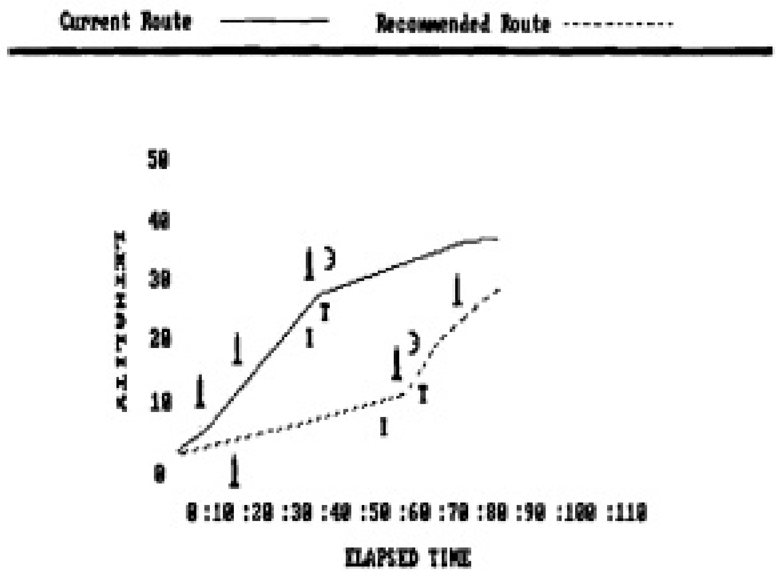

Current Route
Recommended Route
50
40
30
20
10
0
0 :10 :20 :30 :40 :50 :60 :70 :80 :90 :100 :110
ELAPSED TIME

Figure 5-6

DATA: Radar signal at bearing = 183°, A/C at 38° long., 42° lat.
Sources: SAR/Intel Reason: New Threat or Location Est. CONFLICT

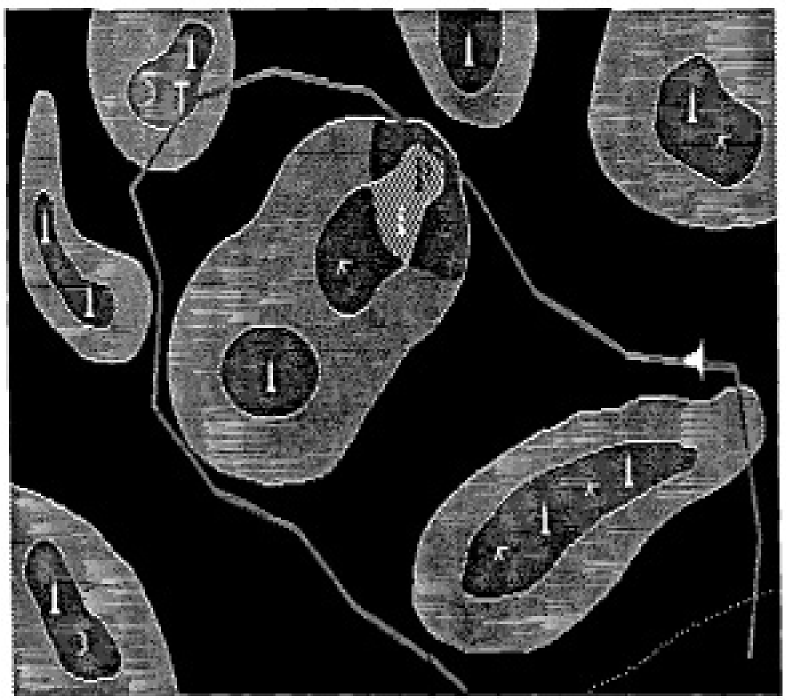


Figure 5-7

inference, conflict resolution, and sensor redeployment have failed thus far to resolve the conflict. In effect, it represents a *test* for the presence of a discount factor which treats the user as a potential source of information. This prompt alerts the user that *he* may possess information which *matters* in the resolution of the conflict. Such prompts do not occur trivially. Their occurrence is determined (as we have seen above) by a comparison of the utility of the test (i.e., the information that may be possessed by the user) with the cost of requesting that information. The estimate of cost should, in principle, be highly sensitive to the current prevailing workload and time stress on the pilot. It should reflect the cost of diverting the pilot from other tasks, as well as the system-computed time remaining until a decision regarding the pop-up threat must be made. In ARR, therefore, the degree of human interaction will vary automatically with circumstances. (Note that even when he has been prompted, the user may decline to respond. In that case, the automated conflict resolution process will resume, for example, by proceeding to a higher cost test or to the phase of overall discounting.)

In the time-stressed cockpit environment, under conditions of conflicting evidence, the CREDIBILITY screen directs the pilot's attention to those assessments (a) about which he is likely to have some information, and (b) which are likely to have the most impact on conflict resolution. Credibility factors which satisfy these criteria are highlighted.

Under conditions of inconsistent evidence, the SITUATION display serves a dual purpose: (1) it prompts the user regarding the conflict (if appropriate), and indicates the nature of the conflict (in yellow) on the spatial display; (2) at the same time, however, its primary purpose remains the display of an aggregated set of danger contours incorporating all relevant information. Thus, it does not provide a vivid or concrete picture of the implications of the conflict to the pilot. Similarly, the RECOMMENDED display provides a route revision which incorporates all currently available information. That is, this is the "compromise" route considered optimal by the aid under the condition that no further information (which might resolve the conflict) were to be obtained. If he chooses however, the user may examine in a "what-if" fashion various ways in which the conflict might be resolved and their implications for route selection. Thus by selecting POSSIBILITIES, he may view alternative conflict resolutions. For example, screen 5-8 shows danger contours which

DATA: Radar signal at bearing = 183^0, A/C at 38^0 long., 42^0 lat.

Source: SAR Assume: New Threat CONFLICT

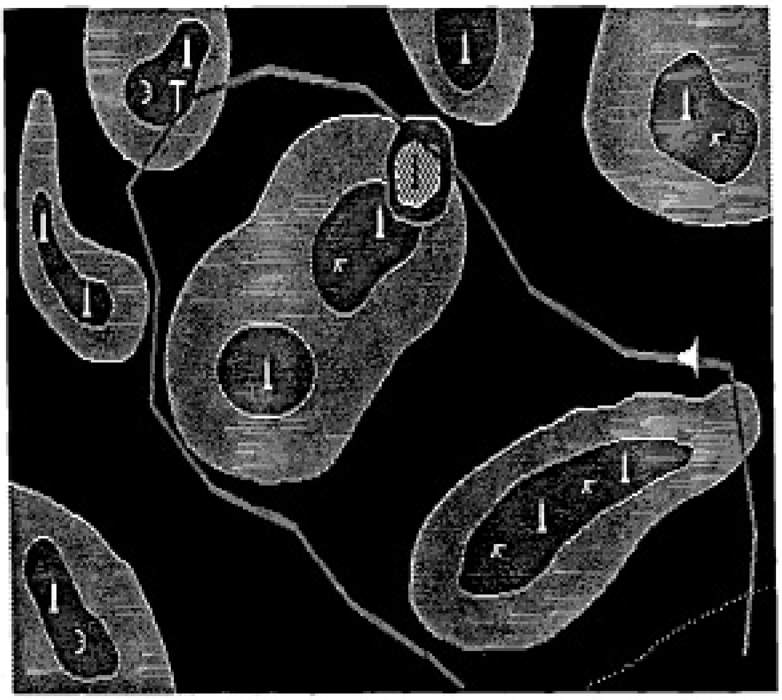


Figure 5-8

would exist if we were to *assume* that the correct interpretation of the new SAR signal involves a previously unidentified threat. Correspondingly, Figure 5-9 shows danger contours which would exist were we to *assume* that the correct interpretation of a SAR signal involves a previously identified threat. If the user now selects RECOMMEND under either of these possibilities, the aid displays the recommended route revision which would be appropriate if the currently selected "possibility" were to be realized (Figures 5-10 and 5-11).

By these means, the user is able to obtain a quick appreciation of the nature of the conflict (i.e., how his mental picture of the tactical situation would change under different conflict resolutions) and the implications of the conflict for his choice of a route. Such a concrete representation of alternative possibilities is a more natural representation of uncertainty in this situation than the more aggregated (but equally necessary) "compromise" displays provided by the SITUATION screen. It corresponds to the pilot's desire to think concretely about "what is out there", and yet at the same time does not permit him to ignore the uncertainty inherent in that process.

The implemented demonstration system focuses primarily on the resolution of conflict regarding the number and localization of threats. However, the inference mechanism and interface design are equally applicable to conflict of evidence in virtually any inferential problem. Figure 5-12 illustrates a SITUATION screen in which conflict regarding the ID of a threat is indicated. In this case, the REASONS screen (Figure 5-13) shows the relative support for various ID possibilities (SA-2, SA-4, etc.) and indicates the reasons which confirm or disconfirm each possibility; similarly, CREDIBILITY indicate factors which influence the credibility of each source of evidence. The POSSIBILITY screens (Figures 5-14 and 5-15) show the implications in terms of danger contours for each possible resolution of the conflict, and RECOMMEND shows the implications for route selection.

A somewhat different, but equally appropriate, use of the present interface design is for in-flight retasking. Figure 5-16 illustrates a SITUATION screen in which the aircraft has been instructed to engage a different target, and the pilot has used RECOMMEND to request ARR to provide a recommended route revision.

DATA: Radar signal at bearing = 183^0, A/C at 38^0 long., 42^0 lat.

Source: SAR   Assume: Location Est.   CONFLICT

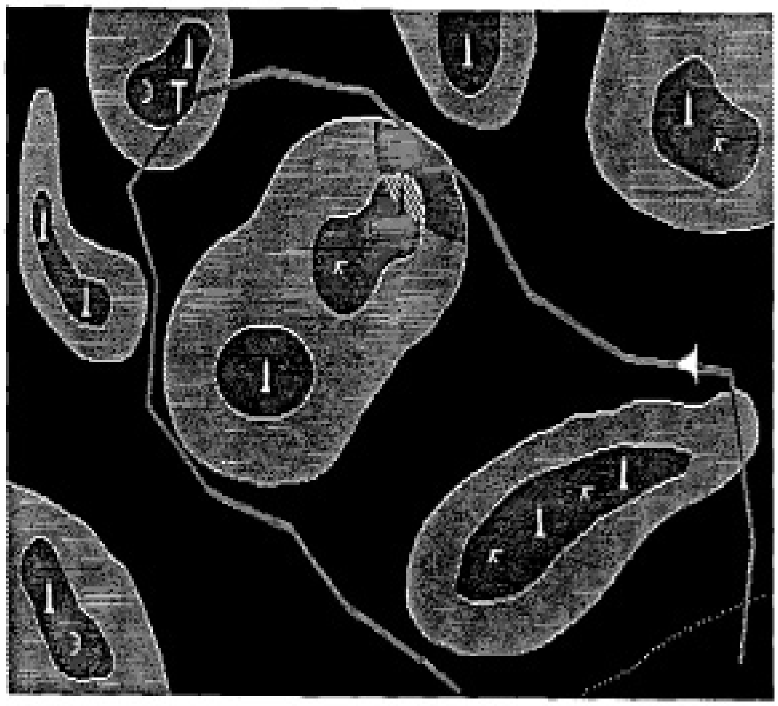


Figure 5-9

DATA: Radar signal at bearing = 183^0, A/C at 38^0 long., 42^0 lat.

Source: SAR Assume: New Threat CONFLICT

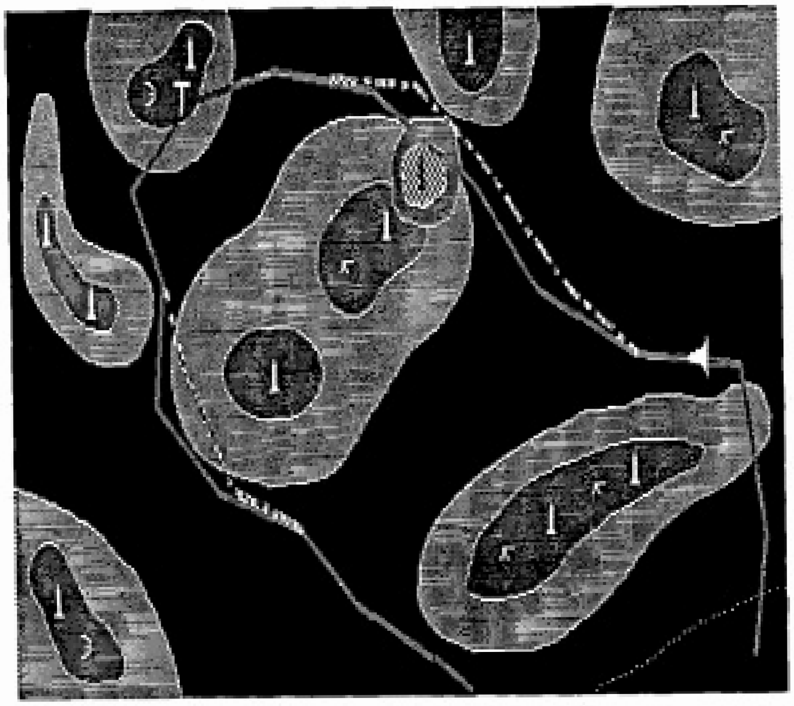


Figure 5-10

DATA: Radar signal at bearing = 183°, A/C at 38° long., 42° lat.

Source: SAR Assume:Threat6 Unchanged CONFLICT

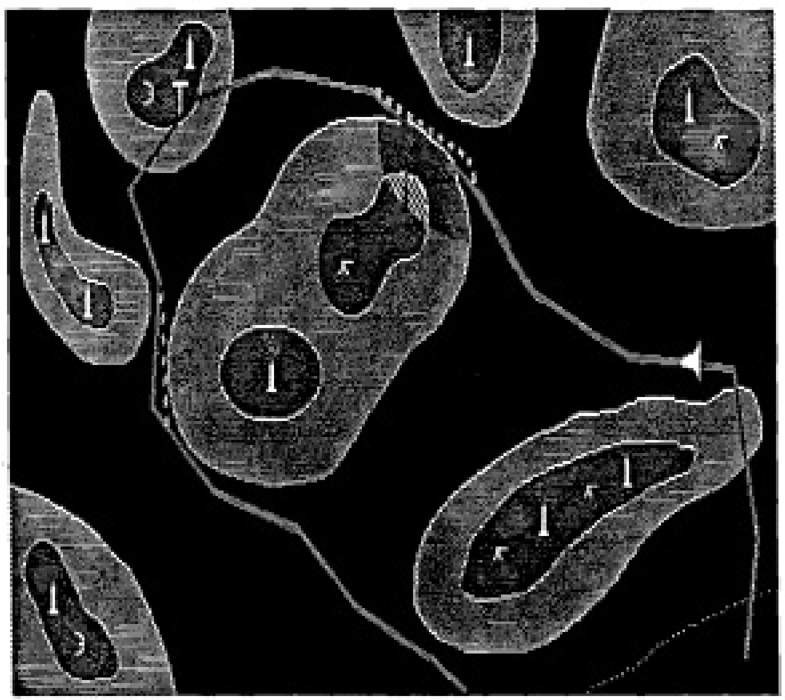


Figure 5-11

DATA: EW Signal

Sources: EW/Intel+Visual   Reason: SA-7 or SA-4   CONFLICT

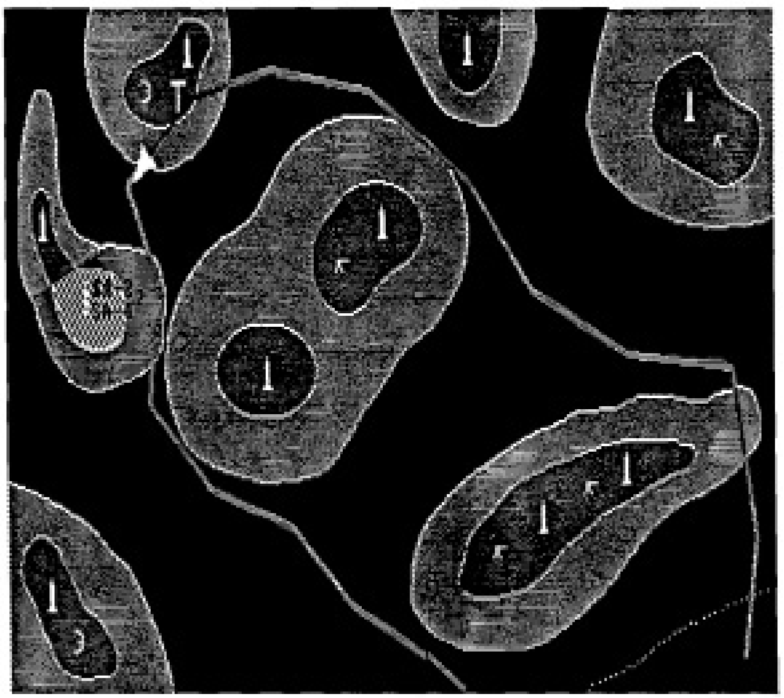


Figure 5-12

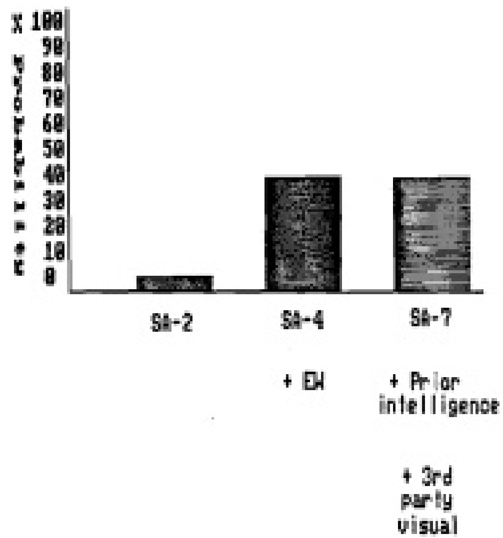

%
100
90
80
70
60
50
40
30
20
10
0
SA-2
SA-4
SA-7
+ EW
+ Prior intelligence
+ 3rd party visual

Figure 5-13

DATA: EW Signal

Sources: EW/Intel*Visual Reason: Assume SA-4 CONFLICT

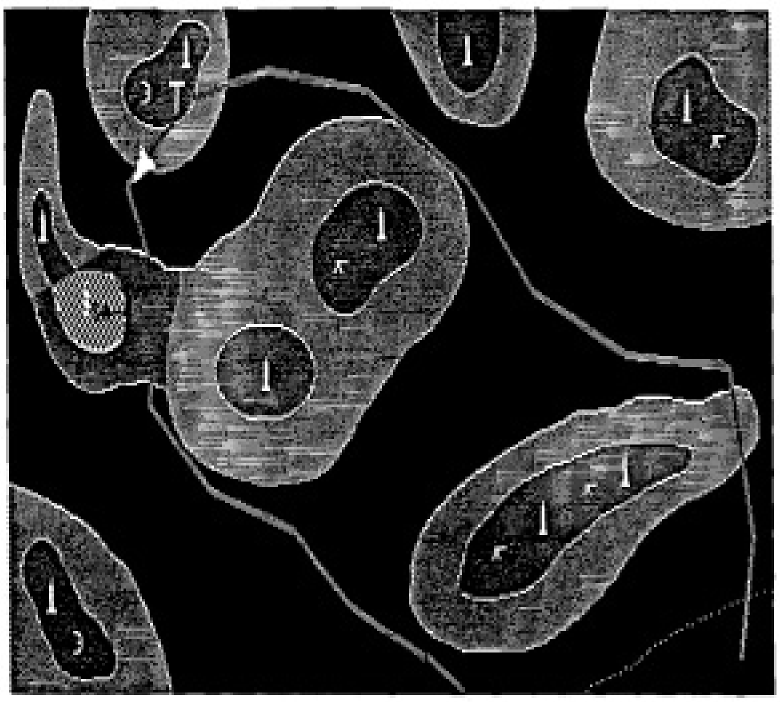


Figure 5-14

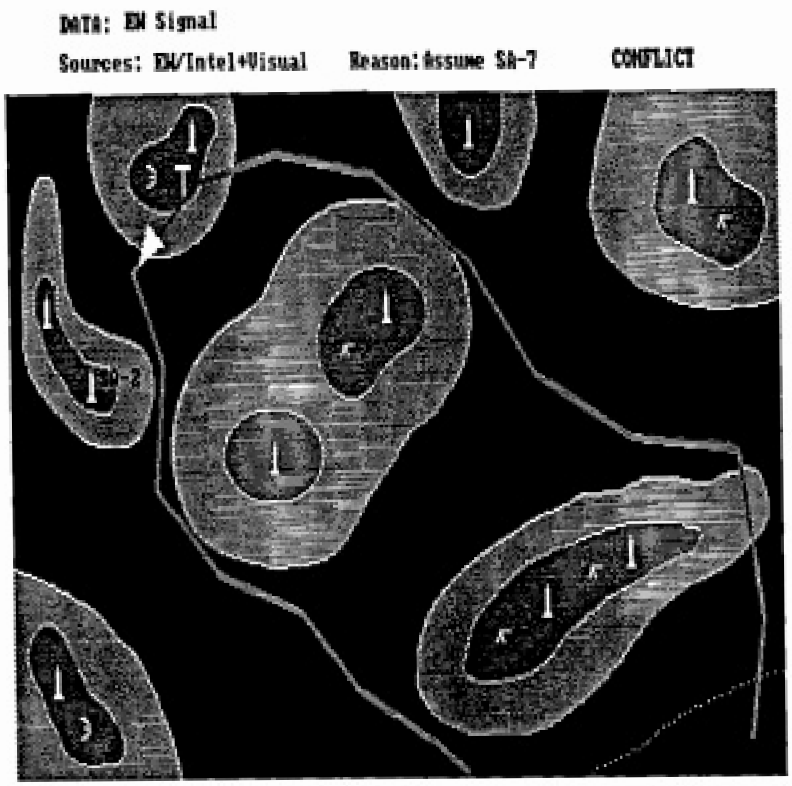

DATA: EW Signal
Sources: EW/Intel+Visual
Reason: Assume SA-7
CONFLICT

Figure 5-15

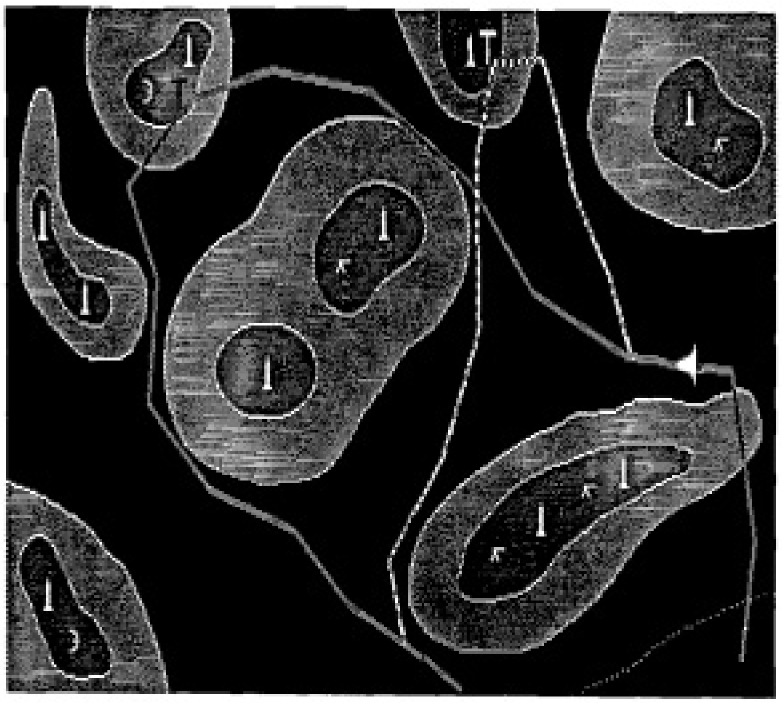


Figure 5-16

6.0 CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

Development of an adaptive self-revising inference engine for handling uncertainty can make a significant contribution to the technology of decision aiding in real-time tactical environments. The research described above has demonstrated the feasibility of such a concept. It has produced a design for expert systems inferencing with very wide applicability. In virtually all problem solving domains where expert systems technology might be introduced, there is need for explicit and valid quantitative modeling of uncertainty. At the same time, there is need for a metastructure of qualitative reasoning in which the assumptions utilized in the probability model are reassessed and revised in the course of the argument. These are the dual requirements addressed by the inference framework described in Section 3 and implemented in the system described in Sections 4 and 5 above.

The next logical step in this research is to go beyond the prototype system described in Sections 4 and 5, to the development, implementation, and testing of a completed system for in-flight route replanning. The resultant system should have immediate relevance to current Air Force efforts to introduce highly promising new technologies into aircraft avionics.

Successful development of such a system would have repercussions going well beyond the specific application of in-flight route replanning. Together with the theoretical framework described above the successful implementation of a completed system would result in the existence of a powerful technology for the building of expert systems in a wide variety of domains. What is learned in this application could be applied in much greater generality, enabling the building of systems capable of accommodating uncertainty both at the level of probabilistic reasoning and at the level of qualitative testing and revising of assumptions.

A completed in-flight route replanning system would require further refinements in the design and algorithms implemented in the prototype system developed during the present effort. Particular developments needed are the exploration of more general forms of discounting, alternative ways of

prioritizing information search, more general sensor management and user interaction capability, as well as other refinements. In all of these refinements, the aim is to implement the inference mechanism in as modular a fashion as possible and as independently as possible from specific domain knowledge. Such an effort woudl result in a generically useful expert systems building tool, suitable for a wide variety of applications domains.

Another crucial feature of expert systems implementation is the incorporation of expert knowledge into the system. Despite its importance, knowledge elicitation continues to be an ill-defined and eclectic art which demands enormous amounts of time from both computer scientists and domain specialists. Another promising avenue for other research, therefore, is an exploration of the implications of our inference framework for knowledge elicitation, both specific to the in-flight route replanning application, and more generally across application domains.

The inference framework developed under this research may contribute in three different ways to progress in automating and streamlining the knowledge elicitation process. First, our framework, by allowing the building of adaptive, self-improving systems, already provides mechanisms for learning and altering the system's reasoning mode in changing environments. Second, by offering a highly structured framework for representing knowledge and manipulating arguments, it provides a type of support for the knowledge elicitation process not afforded by other expert system frameworks. Finally, it may form the eventual basis of an automated knowledge elicitation tool which applies and reconciles multiple methods of eliciting expert knowledge.

REFERENCES

Abelson, R.P. A variance explanation paradox: When a little is a lot. Psychological Bulletin, 1985, 97, 129-133.

Adams, J.B. Probabilistic reasoning and certainty factors. In Buchanan, B.G. and Shortliffe, E.H. (Eds.), *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project*, Reading, MA: Addison-Wesley Publishing Co., 1984, 263-272.

Brown, R.V., and Lindley, D.V. Improving judgment by reconciling incoherence. Theory and Decision, 1982, 14, 113-132.

Brown, R.V., and Lindley, D.V. Plural analysis: Multiple approaches to quantitative research. Theory and Decision, 1985 (in press).

Buchanan, B.G., and Duda, R.O. Principles of rule-based expert systems (Report No. STAN-CS-82-926). Stanford, CA: Stanford University, August 1982.

Buchanan, B.G., and Shortliffe, E.H. Rule-based expert systems. Addison-Wesley, 1984.

Cohen, L.J. The probable and the provable. Oxford, England: Clarendon Press, 1977.

Cohen, M.S., Mavor, A., and Kidd, J. Research on the elicitation of expert knowledge (Proposal). Falls Church, VA: Decision Science Consortium, Inc., March 1984.

Cohen, P.R. Heuristic reasoning about uncertainty: An artificial intelligence approach. Boston, MA: Pitman, 1985.

Cohen, P.R., and Feigenbaum, E.A. (Eds.), The handbook of artificial intelligence. Stanford, CA: HeurisTech Press, 1982, Vol. III.

Cohen, M.S., Schum, D.A., Freeling, A.N.S., and Chinnis, J.O., Jr. On the art and science of hedging a conclusion: Alternative theories of uncertainty in intelligence analysis (Technical Report 84-6). Falls Church, VA: Decision Science Consortium, Inc., August 1985.

de Dombal, F.T. Surgical diagnosis assisted by computer. Proceedings of the Royal Society, 1973, V-184, 433-440.

deFinetti, B. Foresight: Its logical laws, its subjective sources. English translation in H.E. Kybert, Jr., and H.E. Smokler (Eds.), Studies in subjective probability. NY: Wiley, 1964. (Original: 1937)

deGroot, M.H. Comment (On Lindley's paradox). Journal of the American Statistical Association, June 1982, 77(378), 336-339.

Doyle, J. A truth maintenance system. Artificial Intelligence, 1979, 12(3), 231-272.

Dubois, D., and Prade, H. Fuzzy logics and the generalized modus porens revisited (Working Paper). Toulouse, France: Laboratoire Langages et Systemes Informatiques, Universite Paul Sabatier, 1984.

Duda, R., Gaschnig, J., and Hart, P. Model design in the PROSPECTOR consultant system for mineral exploration. In D. Michie (Ed.), Expert systems in the microelectronic age. Edinburgh University Press, 1979, 153-167.

Edwards W. (Ed.). Revisions of opinions by men and man-machine systems. IEEE Transactions on Human Factors in Electronics, 1966, 7(1).

Freeling, A.N.S. Fuzzy sets and decision analysis. IEEE Transactions of Systems, Man, and Cybernetics, 1980, 10(7), 341-354.

Freeling, A.N.S. Possibilities versus fuzzy probabilities--Two alternative decision aids. In H.-J. Zimmerman, and L.A. Zadeh (Eds.), Decision Analysis Through Fuzzy Sets TIMS/ORSA Studies in Management Science, 1983.

Freeling, A.N.S., and Sahlin, N. Combining evidence. In P. Gardenfors, B. Hansson, and N. Sahlin (Eds.), Evidentiary values: Philosophical, judicial, and psychological aspects of a theory. Lund, Sweden: C.W.K. Gleerups, 1983.

Ginsberg, M.L. Non-monotonic reasoning using Dempster's rule. AAAI-84, Austin, TX, 1984, 126-129.

Glymour, C. Theory and evidence. Princeton, NJ: Princeton University Press, 1980.

Gordon, J., and Shortliffe, E.H. The Dempster-Shafer theory of evidence. In B.G. Buchanan, and E.H. Shortliffe (Eds.), Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project, 1984, 272-295.

Hayes-Roth, F., Waterman, D.A., and Lenat, D.B. Building expert systems. Reading: MA: Addison-Wesley Publishing Co., Inc., 1983.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.). Judgment under uncertainty: Heuristics and biases. NY: Cambridge University Press, 1982.

Kim, J.H. CONVINCE: A conversational inference consolidation engine (Doctoral Dissertation). Los Angeles, CA: University of California, 1983.

Levi, I. Consonance, dissonance and evidentiary mechanisms. In P. Gardenfors, B. Hansson, and N. Sahlin (Eds.), Evidentiary value: Philosophical, judicial and psychological aspects of a theory, Lund, Sweden: C.W.K. Gleerups, 1983.

Lindley, D.V. Scoring rules and the inevitability of probability. International Statistical Review, 1982, 50, 1-26.

Lindley, D.V. Reconciliation of discrete probability distributions. In J.M. Bernardo et al. (Eds.), Proceedings of the 2nd Valencia International Conference on Bayesian Statistics, 1984.

Lindley, D.V., Tversky, A., and Brown, R.V. On the reconciliation of probability assessments. Journal of the Royal Statistical Society., Series A, 1979, 142(2), 146-180.

Lindsay, R., Buchanan, B.G., Feigenbaum, E.A., and Lederberg, J. Applications of artificial intelligence for organic chemistry: DENDRAL. NY: McGraw-Hill, 1980.

McDermott, D., and Doyle, J. Non-monotonic logic I. Artificial Intelligence, 1980, 13, 41-72.

Nilsson, N.J. Probabilistic logic (Technical Note 321). Menlo Park, CA: SRI International, February 1984.

Pearl, J. Distributed Bayesian belief maintenance. Proceedings of the Second National Conference on Artificial Intelligence. Los Altos, CA: William Kaufmann, Inc., 1982.

Quine, W.V. Two dogmas of empiricism. In W.V. Quine (Ed.), From a logical point of view, NY: Harper and Row, Inc., 1953.

Reiter, R. A logic for default reasoning. Artificial Intelligence, 1980, 13, 81-132.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10). NY: Academic Press, 1977, 174-177.

Savage, L.J. The foundations of statistics. New York: Wiley, 1954.

Schum, D.A. A review of a case against Blaise Pascal and his heirs. Michigan Law Review, Jan-Mar 1979, 77(3), 446-484.

Schum, D.A. Current developments in research on cascaded inference processes. Chapter 10 of T.S. Wallsten (Ed.), Cognitive process in choice and decision behavior. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Schum, D.A. Sorting out the effects of witness sensitivity and response criterion placement upon the inferential value of testimonial evidence. Organizational Behavior and Human Performance, 1981, 2.

Schum, D.A., and Martin, A.W. Probabilistic opinion revision on the basis of evidence at trial: A Baconian or a Pascalian process? (Report 80-02). Houston, TX: Rice University, 1980.

Shafer, G. A mathematical theory of evidence. Princeton, NJ: Princeton University Press, 1976.

Shafer, G. Jeffrey's rule of conditioning. Phil. of Sci., 1981, 48, 337-362.

Shafer, G. Probability judgment in artificial intelligence and expert systems. Lawrence, KS: University of Kansas, School of business, December 1984.

Shafer, G. Belief functions and possibility measures. Lawrence, KS: University of Kansas, School of Business, in press.

Shafer, G., and Tversky, A. Weighing evidence: The design and comparison of probability thought experiments. Stanford, CA: Stanford University, June 1983.

Shortliffe, E.H. Computer based medical consultation: MYCIN. Elsevier, 1976.

Slovic, P., and Tversky, A. Who accepts Savage's axiom? Behavioral Science, 1974, 19, 368-373.

Stallman, R.M., and Sussman, G.J. Problem solving about electrical circuits. In Proceedings of the Fifth International Joint Conference on Artificial Intelligence, August 22-25, 1977. Cambridge, MA: 299-304.

Tversky, A. and Kahneman, D. Causal schemas in judgments under uncertainty. In D. Kahneman, P. Slovic, and A. Tversky, Judgment under uncertainty: Heuristics and biases, Cambridge University Press, 1982, 117-128.

Tversky, A., and Kahneman, D. Extensional vs. intuitive reasoning: The conjunctive fallacy in probability judgment. Psychological Review, 1983, 90, 193-315.

Watson, S.R., Weiss, J.J., and Donell, M.L. Fuzzy decision analysis. IEEE Transactions of Systems, Man and Cybernetics, 1979, SMC-9, 1-9.

Williams, P.M. On a new theory of epistemic probability. The British Journal for the Philosophy of Science, 1978, 29, 375-387.

Zadeh, L.A. Fuzzy sets. Inf. and Contr., 1965, 8, 338-353.

Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning. Information Science, 1975, 8, 199-249; 301-357; 9, 43-80.

Zadeh, L.A. Possibility theory and soft data analysis. In L. Cobb, and R.M. Thrall (Eds.), Mathematical frontiers of the social and policy sciences. AAAS: Washington, 1981, 69-129.

Zadeh, L.A. Review of Shafer's, A mathematical theory of evidence. AI Magazine, 1984, 5(3), 81-83.